Certificate of Mailing

Date of Deposit January 26, 2000

Label Number: EL509044753US

I hereby certify under 37 CFR 1.10 that this correspondence is being deposited with the United States Postal Service as **"Express Mail Post Office to Addressee"** with sufficient postage on the date indicated above and is addressed to BOX PATENT APPLICATION, Assistant Commissioner of Patents, Washington, D.C. 20231.

LUIS CRUZ

Printed name of person mailing correspondence

Signature of person mailing correspondence

Substitute Form PTO 1390 U.S. Department of Commerce Patent and Trademark Office TRANSMITTAL LETTER TO THE UNITED STATES DESIGNATED/ELECTED OFFICE (DO/EO/US) CONCERNING A FILING UNDER 35 U.S.C. 371	Attorney's Docket Number: 07005/003002 U.S. Application Number: Not yet assigned

INTERNATIONAL APPLICATION NUMBER	INTERNATIONAL FILING DATE	PRIORITY DATE CLAIMED
PCT/US98/15563	28 July 1998	29 July 1997

TITLE OF INVENTION:	PREDICTION, DETECTION, AND DESIGN OF T CELL EPITOPES
APPLICANTS FOR DO/EO/US:	Samuel J. Landry

Applicant herewith submits to the United States Designated/Elected Office (DO/EO/US) the following items and other information:

1.	X	This is a FIRST submission of items concerning a filing under 35 U.S.C. 371.
2.		This is a SECOND or SUBSEQUENT submission of items concerning a filing under 35 U.S.C. 371.
3.	X	This is an express request to begin national examination procedures (35 U.S.C. 371(f)) at any time rather than delay examination until the expiration of the applicable time limit set in 35 U.S.C. 371(b) and PCT Articles 22 and 39(1).
4.	X	A proper Demand for International Preliminary Examination was made by the 19th month from the earliest claimed priority date.
5.		A copy of the International Application as filed (35 U.S.C. 371(c)(2)).
a.	X	is transmitted herewith (required only if not transmitted by the International Bureau).
b.		has been transmitted by the International Bureau.
c.		Is not required, as the application was filed with the United States Receiving Office (RO/US).
6.		A translation of the International Application into English (35 U.S.C. 371(c)(2).
7.		Amendments to the claims of the International Application under PCT Article 19 (35 U.S.C. 371(c)(3)).
a.		are transmitted herewith (required only if not transmitted by the International Bureau).
b.		have been transmitted by the International Bureau.
c.		have not been made; however, the time limit for making such amendments has NOT expired.
d.	X	have not been made and will not be made.
8.		A translation of the amendments to the claims under PCT Article 19 (35 U.S.C. 371(c)(3)).
9.	X	An oath or declaration of the inventors (35 U.S.C. 371(c)(4)). (UNSIGNED)
10.		A translation of the annexes to the International Preliminary Examination Report under PCT Article 36 (35 U.S.C. 371 (c)(5).
11.		An Information Disclosure Statement under 37 CFR 1.97 and 1.98.

12.		An assignment for recording. A separate cover sheet in compliance with 37 3.28 and 3.31 is included.				
13.	X	A FIRST preliminary amendment.				
		A SECOND or SUBSEQUENT preliminary amendment.				
14.		A substitute specification.				
15.		A change of power of attorney and/or address letter.				
16.		Other items or information:				
17.		The following fees are submitted: BASIC NATIONAL FEE (37 CFR 1.492(A)(1)-(5)): International preliminary examination fee paid to USPTO (37 CFR 1.482) $ 670.00				
				ENTER APPROPRIATE BASIC FEE AMOUNT =	$ 670.00	
		Surcharge of $130 for furnishing the oath or declaration later than □ 20 OR □ 30 months from the earliest claimed priority date (37 CFR 1.492(e)).			$	
CLAIMS		NUMBER FILED	NUMBER EXTRA	RATE		
Total claims		41 - 20 =	21	x $22.00	$ 462.00	
Independent claims		3 - 3 =	0	x $82.00	$	
		Multiple dependent claims (if applicable)		+ $270.00	$ 270.00	
				TOTAL OF ABOVE CALCULATIONS =	$ 1402.00	
		Reduction of ½ for filing by small entity, if applicable. Verified Small Entity Statement must also be filed with this request (Note 37 CFR 1.9, 1.27, 1.28).			$ 701.00	
				SUBTOTAL =	$ 701.00	
		Processing fee of $130.00 for furnishing the English translation later than □ 20 OR □ 30 months from the earliest claimed priority date (37 CFR 1.492(f)). +			$	
				TOTAL NATIONAL FEE =	$	
		Fee for recording the enclosed assignment (37 CFR 1.21(h)). The assignment must be accompanied by an appropriate cover sheet (37 CFR 3.28, 3.31). $40.00 per property. +			$	
				TOTAL FEES ENCLOSED =	$ 701.00	
					Amount to be refunded	$
					charged	$

a.	X	A check in the amount of $ 701.00 to cover the above fees is enclosed.
b.		Please charge my Deposit Account No. 03-2095 in the amount of $ [**.**] to cover the above fees.
c.	X	The Commissioner is hereby authorized to charge any additional fees which may be required, or credit any overpayment, to Deposit Account No. 03-2095. A duplicate copy of this sheet is enclosed.

NOTE: Where an appropriate time limit under 37 CFR 1.494 or 1.495 has not been met, a petition to revive (37 CFR 1.137(a) or (b) must be filed and granted to restore the application to pending status.

SEND ALL CORRESPONDENCE TO:	
Kristina Bieker-Brady, Ph.D. Clark & Elbing LLP 176 Federal Street Boston, MA 02110-2214 Telephone: 617-428-0200 Facsimile: 617-428-7045	Signature Kristina Bieker-Brady, Ph.D. Reg No. 39,109

07005.003002 US Utility (National Phase).wpd

PATENT

ATTORNEY DOCKET NO. 07005/003002

Certificate of Mailing

Date of Deposit January 26, 2000 Label Number: EL509044753US

I hereby certify under 37 CFR 1.10 that this correspondence is being deposited with the United States Postal Service as **"Express Mail Post Office to Addressee"** with sufficient postage on the date indicated above and is addressed to BOX PATENT APPLICATION, Assistant Commissioner of Patents, Washington, D.C. 20231.

LUIS CRUZ

Printed name of person mailing correspondence

Signature of person mailing correspondence

IN THE UNITED STATES PATENT AND TRADEMARK OFFICE
IN THE UNITED STATES RECEIVING OFFICE (US/RO)

Applicant : SAMUEL J. LANDRY
Serial No.: Not yet assigned; based on PCT/US98/15563
Filed : Herewith
Title : PREDICTION, DETECTION, AND DESIGN OF T CELL EPITOPES

Box PCT
Assistant Commissioner for Patents
Washington, D.C. 20231

PRELIMINARY AMENDMENT

In connection with the attached application, Applicant makes the following amendments.

In the Claims

Please cancel claims 20-44.

If there are any charges, or any credits, please apply them to Deposit Account No. 03-2095.

Respectfully submitted,

Date: January 24, 2000

Kristina Bieker-Brady, Ph.D.
Reg. No. 39,109

Clark & Elbing LLP
176 Federal Street
Boston, MA 02110
Telephone: 617-428-0200
Facsimile: 617-428-7045

07005.003002 Preliminary Amendment.wpd

ATTORNEY DOCKET NO. 07005/003002

Applicant or Patentee : Samuel J. Landry
Serial or Patent No. :
Filed or Issued : Herewith
Title : PREDICTION, DETECTION, AND DESIGN OF T-CELL EPITOPES

VERIFIED STATEMENT (DECLARATION) CLAIMING SMALL ENTITY STATUS (37 CFR 1.9(f) and 1.27(d)) - NONPROFIT ORGANIZATION

I hereby declare that I am an official empowered to act on behalf of the nonprofit organization identified below:

Name of Organization: The Administrators of the Tulane Educational Fund
Address of Organization: 1430 Tulane Avenue, New Orleans, Louisiana 70112-2699
Type of Organization:

[X] University or Other Institution of Higher Education
[ ] Tax Exempt under Internal Revenue Service Code (26 USC 501(a) and 501(c)(3))
[ ] Nonprofit Scientific or Educational under Statute of State of the United States of America
Name of State:
Citation of Statute:
[ ] Would Qualify as Tax Exempt under Internal Revenue Service Code (26 Usc 501(a) and 501(c)(3)) If Located in the United States of America
[ ] Would Qualify as Nonprofit Scientific or Educational under Statute of State of the United States of America If Located in the United States of America
Name of State:
Citation of Statute:

I hereby declare that the nonprofit organization identified above qualifies as a nonprofit organization as defined in 37 CFR 1.9(e) for purposes of paying reduced fees under section 41(a) and (b) of Title 35, United States Code with regard to the invention entitled PREDICTION, DETECTION, AND DESIGN OF T-CELL EPITOPES by Inventors Samuel J. Landry described in

[X] the specification filed herewith.
[ ] application serial no. _____, filed _____.
[ ] patent no. ["PATENT NUMBER"], issued ["ISSUE DATE"].

I hereby declare that rights under contract or law have been conveyed to and remain with the nonprofit organization with regard to the above identified invention.

If the rights held by the nonprofit organization are not exclusive, each individual, concern or organization having rights to the invention is listed below* and no rights to the invention are held by any person, other than the inventor, who could not qualify as a small business concern under 37 CFR 1.9(c) or by any concern which would not qualify as a small business concern under 37 CFR 1.9(d) or a nonprofit organization under 37 CFR 1.9(e).

*NOTE: Separate verified statements are required from each named person, concern or organization having rights to the invention averring to their status as small entities. (37 CFR 1.27)

Full Name:
Address:
[ ] INDIVIDUAL [ ] SMALL BUSINESS CONCERN [ ] NONPROFIT ORGANIZATION

I acknowledge the duty to file, in this application or patent, notification of any change in status resulting in loss of entitlement to small entity status prior to paying, or at the time of paying, the earliest of the issue fee or any maintenance fee due after the date on which status as a small entity is no longer appropriate. (37 CFR 1.28(b))

I hereby declare that all statements made herein of my own knowledge are true and that all statements made on information and belief are believed to be true; and further that these statements were made with the knowledge that willful false statements and the like so made are punishable by fine or imprisonment, or both, under section 1001 of Title 18 of the United States Code, and that such willful false statements may jeopardize the validity of the application, any patent issuing thereon, or any patent to which this verified statement is directed.

Name: Carla Fishman

Title: Executive Director, Office of Technology Development

Address: Tulane University Medical Center, 1430 Tulane Avenue, New Orleans, Louisiana 70112-2699

Signature: Carla N Fish **Date:** 1/26/00

PREDICTION, DETECTION, AND DESIGN OF T CELL EPITOPES

Background of the Invention

Antigen-specific immunity is mediated by lymphocytes, a specialized group of white blood cells. The two types of lymphocytes, B and T, are capable of recognizing and binding to an antigen. B lymphocytes typically recognize and bind to an entire antigen (*i.e.*, a long polypeptide chain that has both secondary and tertiary structural elements). On the other hand, T lymphocytes bind to a peptide that is derived from the entire antigen in a mechanism termed antigen processing. T lymphocytes will only recognize the peptide when it is presented in context with major histocompatibility complex (MHC) proteins expressed on the surface of another cell.

There are two classes of MHC molecules, class I and class II, which are recognized by $CD8^+$ and $CD4^+$ T lymphocytes, respectively. Very different antigen processing mechanisms generate the majority of peptides found in class I and class II major histocompatibility (MHC) proteins (Germain and Margulies, Ann. Rev. Immunol. 11: 403-450, 1993; Germain, G., Cell 76: 287-299, 1994). Mechanisms of antigen processing and presentation limit the spectrum of peptides that can be recognized by T-cell receptors. Nevertheless, for a given antigen sequence, many more peptides should satisfy the requirements of MHC binding than actually are observed in the population of naturally processed and presented peptides. The peptides actually observed are called the immunodominant epitopes of the antigen.

The ability to selectively predict and generate immunodominant epitopes would improve vaccine development and therapies for immunological diseases.

Summary of the Invention

I have discovered a method which allows detection and creation of immunodominant T cell epitopes. This method allows the generation of improved methods for treating diseases involving immune responses mediated by antigen-

specific T lymphocytes.

In a first aspect, the invention features a method for stimulating an immune response specific toward a naturally-occurring protein in an animal having an immune system that includes T cells, where the method includes administering to the animal an altered protein derived from the naturally-occurring protein, wherein an unstable polypeptide segment has been inserted by artifice into the altered protein. In one embodiment of the first aspect of the invention, the naturally occurring-protein is from a pathogen, and the altered protein is administered to the animal to prevent infection of the animal with the pathogen. In another embodiment, the naturally occurring-protein is from a neoplastic cell, and the altered protein is administered to the animal to inhibit growth of the neoplastic cell in the animal. Preferably, the altered protein is administered with a pharmaceutically acceptable carrier, an adjuvant, or both. Preferably, the animal is a mammal (*e.g.*, a human).

In a second aspect, the invention provides a method for increasing the immunogenicity of a naturally-occurring protein by inserting by artifice into the naturally-occurring protein an unstable polypeptide segment to produce an altered protein. In one embodiment of this aspect of the invention, the naturally-occurring protein is from a pathogen or a neoplastic cell. Preferably, the altered protein, or polypeptide fragment thereof, is in a vaccine.

In one embodiment of the first and second aspects of the invention, the unstable polypeptide segment includes at least twelve amino acid residues. In another embodiment, not more than 30% of the amino acid residues of the unstable polypeptide segment are isoleucine, leucine, valine, tyrosine, phenylalanine, tryptophan, threonine, or methionine. In another embodiment, the unstable polypeptide segment includes a polypeptide sequence that is specifically recognized by a protease. In other embodiments, the unstable polypeptide segment has an average hydrophobicity value that is lower than the average hydrophobicity value of the altered protein; has a sequence conservation that is lower than a sequence

conservation of the altered protein; has an amide protection factor that is lower than 10^4 wherein the altered protein is in a native conformational state; has an average amide protection factor that is lower than the average amide protection factor for the altered protein in a denatured conformational state; has an NMR order parameter (S^2) of less than 0.8, and preferably less than 0.7; or has an average B-factor value that is higher than the average B-factor value of the altered protein.

In a preferred embodiment of the first and second aspects of the invention, the altered protein includes a T cell epitope. In another embodiment, the unstable polypeptide segment is inserted N-terminally adjacent to the T cell epitope. In another embodiment, the C - terminal portion of the unstable polypeptide segment overlaps the N - terminal portion of the T cell epitope. In various other embodiments, the T cell epitope has an average hydrophobicity value that is higher than the average hydrophobicity value of the altered protein; has a sequence conservation that is higher than a sequence conservation of the altered protein; has an amide protection factor that is greater than 10^4 wherein the altered protein is in a native conformational state; has an average amide protection factor that is higher than the average amide protection factor for the altered protein in a denatured conformational state; has an NMR order parameter (S^2) of greater than 0.7; or has an average B-factor value that is lower than the average B-factor value of the altered protein. In another embodiment, at least 30% of the amino acid residues of the T cell epitope are selected from the group of amino acid residues consisting of isoleucine, leucine, valine, tyrosine, phenylalanine, tryptophan, threonine, and methionine.

In a third aspect, the invention features a method for detecting in a substantially pure protein a polypeptide segment that is likely to be a T-cell epitope includes first, identifying an unstable polypeptide segment in the protein; and, second, identifying a second polypeptide segment adjacent to the unstable polypeptide segment, the second polypeptide segment likely to be a T cell epitope.

In one embodiment of the third aspect of the invention, the unstable polypeptide segment includes at least twelve amino acid residues. In another

embodiment, not more than 30% of the amino acid residues of the unstable polypeptide segment are isoleucine, leucine, valine, tyrosine, phenylalanine, tryptophan, threonine, or methionine. In another embodiment, the unstable polypeptide segment includes a polypeptide sequence that is specifically recognized by a protease. In other embodiments, the unstable polypeptide segment has an average hydrophobicity value that is lower than the average hydrophobicity value of the substantially pure protein; has a sequence conservation that is lower than a sequence conservation of the substantially pure protein; has an amide protection factor that is lower than 10^4 wherein the substantially pure protein is in a native conformational state; has an average amide protection factor that is lower than the average amide protection factor for the substantially pure protein in a denatured conformational state; has an NMR order parameter (S^2) of less than 0.8; or has an average B-factor value that is higher than the average B-factor value of the substantially pure protein.

In a preferred embodiment of the third aspect of the invention, the unstable polypeptide segment is N-terminally adjacent to the second polypeptide segment. In another embodiment, the C - terminal portion of the unstable polypeptide segment overlaps the N - terminal portion of the second polypeptide segment. In other embodiments, the second polypeptide segment has an average hydrophobicity value that is higher than the average hydrophobicity value of the protein; has a sequence conservation that is higher than a sequence conservation of the protein; has an amide protection factor that is greater than 10^4 wherein the protein is in a native conformational state; has an average amide protection factor that is higher than the average amide protection factor for the protein in a denatured conformational state; has an NMR order parameter (S^2) of greater than 0.7, and preferably greater than 0.8; or has an average B-factor value that is lower than the average B-factor value of the protein. In yet another embodiment, at least 30% of the amino acid residues of the second polypeptide segment are selected from the group of amino acid residues consisting of isoleucine, leucine, valine, tyrosine, phenylalanine, tryptophan,

threonine, and methionine.

In a fourth aspect, the invention features a method for identifying the most immunogenic protein in a group of proteins, the method including identifying the protein including the most unstable polypeptide segment in the group of proteins, wherein the identified protein is the most immunogenic protein in the group of proteins. Preferably, the most immunogenic protein is substantially purified and the group of proteins is from a neoplastic cell, a pathogen, a foodstuff, an allergen, or a tissue targeted in an autoimmune disease.

In one embodiment of the fourth aspect of the invention, the most unstable polypeptide segment includes at least twelve amino acid residues. In another embodiment, not more than 30% of the amino acid residues of the most unstable polypeptide segment are selected from the group of amino acid residues consisting of isoleucine, leucine, valine, tyrosine, phenylalanine, tryptophan, threonine, and methionine. In various other embodiments, the most unstable polypeptide segment has the lowest average hydrophobicity value of any unstable polypeptide segment of the group of proteins; has the lowest sequence conservation of any unstable polypeptide segment of the group of proteins; has the lowest average amide protection factor of any unstable polypeptide segment of the group of proteins wherein the proteins in the group are in a native conformational state; has the lowest average amide protection factor of any unstable polypeptide segment of the proteins wherein the proteins in the group are in a denatured conformational state; has the lowest NMR order parameter (S^2) of any unstable polypeptide segment of the group of proteins; or has the average highest B-factor value of any unstable polypeptide segment of the group of proteins.

In a preferred embodiment of the fourth aspect of the invention, the most immunogenic protein includes a T cell epitope. In one embodiment, the most unstable polypeptide segment is N-terminally adjacent to the T cell epitope. In another embodiment, the C - terminal portion of the most unstable polypeptide segment overlaps the N - terminal portion of the T cell epitope. In various other

embodiments, the T cell epitope has an average hydrophobicity value that is higher than the average hydrophobicity value of the protein; has a sequence conservation that is higher than a sequence conservation of the protein; has an amide protection factor that is greater than 10^4 wherein the protein is in a native conformational state; has an average amide protection factor that is higher than the average amide protection factor for the protein in a denatured conformational state; has an NMR order parameter (S^2) of greater than 0.7; or has an average B-factor value that is lower than the average B-factor value of the protein. In yet another embodiment, at least 30% of the amino acid residues of the T cell epitope are selected from the group of amino acid residues consisting of isoleucine, leucine, valine, tyrosine, phenylalanine, tryptophan, threonine, and methionine.

In a fifth aspect, the invention features a method for treating an animal that has or is at risk for developing an allergic response, the method including administering to the animal a protein or polypeptide fragment thereof from an allergen, wherein said protein or polypeptide fragment thereof is identified as comprising the most unstable polypeptide segment in a group of proteins of said allergen.

In a sixth aspect, the invention features a method for treating an animal that has or is at risk for developing an autoimmune disease, the method including administering to the animal a protein or polypeptide fragment thereof from a tissue targeted in said immune disease wherein said protein or polypeptide fragment thereof is identified as comprising the most unstable polypeptide segment in a group of proteins of said tissue targeted in said autoimmune disease.

In one embodiment of the fifth and sixth aspects of the invention, the protein or polypeptide fragment thereof is in a tolerogen. In another embodiment, the protein or polypeptide fragment thereof is administered orally, or is administered with a pharmaceutically acceptable carrier. In other embodiments, the animal is a mammal (*e.g.*, a human).

In an seventh aspect, the invention features a substantially pure antigen that includes an unstable polypeptide segment that has been inserted by artifice.

Preferably, the antigen is associated with a pharmaceutically acceptable carrier, an adjuvant, or both.

In one embodiment of the seventh aspect of the invention, the unstable polypeptide segment includes at least twelve amino acid residues. In another embodiment, not more than 30% of the amino acid residues of the unstable polypeptide segment are selected from the group of amino acid residues consisting of isoleucine, leucine, valine, tyrosine, phenylalanine, tryptophan, threonine, and methionine. In another embodiment, the unstable polypeptide segment includes a polypeptide sequence that is specifically recognized by a protease. In various other embodiments, the unstable polypeptide segment has an average hydrophobicity value that is lower than the average hydrophobicity value of the substantially pure antigen; has a sequence conservation that is lower than a sequence conservation of the substantially pure antigen; has an amide protection factor that is lower than 10^4 wherein the substantially pure antigen is in a native conformational state; has an average amide protection factor that is lower than the average amide protection factor for the substantially pure antigen in a denatured conformational state; has an NMR order parameter (S^2) of less than 0.8, and preferably less than 0.7; or has an average B-factor value that is higher than the average B-factor value of the substantially pure antigen.

In a preferred embodiment of the seventh aspect of the invention, the substantially pure antigen includes a T cell epitope. In another embodiment, the unstable polypeptide segment is inserted N-terminally adjacent to the T cell epitope. In yet another embodiment, the C - terminal portion of the unstable polypeptide segment overlaps the N - terminal portion of the T cell epitope. In various other embodiments, the T cell epitope has an average hydrophobicity value that is higher than the average hydrophobicity value of the antigen; has a sequence conservation that is higher than a sequence conservation of the antigen; has an amide protection factor that is greater than 10^4 wherein the antigen is in a native conformational state; has an average amide protection factor that is higher than the average amide

protection factor for the antigen in a denatured conformational state; has an NMR order parameter (S^2) of greater than 0.7; or has an average B-factor value that is lower than the average B-factor value of the antigen. In yet another embodiment, at least 30% of the amino acid residues of the T cell epitope are selected from the group of amino acid residues consisting of isoleucine, leucine, valine, tyrosine, phenylalanine, tryptophan, threonine, and methionine.

In a eighth aspect, the invention features a vaccine that includes the antigen of the seventh aspect of the invention.

In a ninth aspect, the invention features a tolerogen that includes the antigen of the seventh aspect of the invention.

In an tenth aspect, the invention features a method for determining the capacity of an animal to mount an antigen-specific T lymphocyte response that includes first, administering to the animal an antigen identified as having at least one unstable polypeptide segment, second, providing the antigen, or a polypeptide fragment thereof, to a T lymphocyte isolated from the animal, and, third, measuring the response of the T lympchoyte, where a response indicates that the animal is capable of an antigen-specific T lymphocyte response to the antigen.

In an eleventh aspect, the invention features a method for detecting an autoimmune disease, or a predisposition to develop an autoimmune disease, in an animal that includes first, providing an antigen from the animal, or a polypeptide fragment thereof, to a T lymphocyte isolated from the animal, and, second, measuring the response of the T lympchoyte, where a response indicates the animal has or is predisposed to develop an autoimmune disease.

In various embodiments of the above aspects of the invention, a mammal may be a goat, rabbit, pig, horse, donkey, elephant, camel, sheep, guinea pig, rat, mouse, chimpanzee, gorilla, gibbon, llama, macaque, or marine mammal. Animals of the invention may be warm-blooded mammals and non-mammals (*e.g.*, birds), as well as cold-blooded vertebrates and invertebrates. Pathogens of the invention include, without limitation, viruses (*e.g.*, the human immunodeficiency virus), bacteria,

protozoans, helminths, and yeasts.

In a twelfth aspect, the invention features a method for using engineered T cell epitopes to stimulate either TH1 or TH2 helper T lymphocyte responses.

In a thirteenth aspect, the invention features a method for using engineered T cell epitopes to tolerize a patient to an antigen having the T cell epitope by selectively removing epitope-reactive T cells from the patient's blood.

In a fourteenth aspect, the invention features a method for using engineered T cell epitopes to tolerize a patient to an antigen having the T cell epitope by selectively rendering epitope-reactive T cells in the patient's blood anergic by stimulating the T cells with the epitope plus autologous MHC in the absence of co-stimulatory molecules.

By "antigen-presenting cell" is meant a cell that expresses MHC protein on its cell surface. A preferable antigen-presenting cell expresses MHC class II proteins on its cell surface, and a most preferable antigen-presenting cell expresses both MHC class I and MHC class II proteins on its cell surface.

By "epitope" or "T cell epitope" is meant a polypeptide, or a fragment thereof, derived from or corresponding to an antigen that is expressed on the surface of an antigen-presenting cell in context with MHC class I or MHC class II proteins. The cell-bound complex of epitope plus MHC is specifically recognized by the T cell receptor on the cell surface of a T lymphocyte. A T cell epitope is generally a stable polypeptide fragment, with at least 30% of its amino acid residues being isoleucine, leucine, valine, tyrosine, phenylalanine, tryptophan, threonine, or methionine residues, more preferably, at least 40%, still more preferably, at least 50% and most preferably, at least 70% of its amino acid residues being isoleucine, leucine, valine, tyrosine, phenylalanine, tryptophan, threonine, or methionine residues. An T cell epitope may be defined by a high NMR order parameter (S^2). A high NMR order parameter is an NMR order parameter which is greater than 0.7, and, more preferably, greater than 0.8. By "high NMR order parameter" is meant the high degree of order

exhibited by NMR-active nuclei on the time scale of picoseconds to nanoseconds as determined by NMR relaxation processes. Preferably, T cell epitope may be defined as having a higher than average hydrophobicity (as determined by a Kyte-Doolittle scale) relative to the hydrophobicity of the entire polypeptide. More preferably, a T cell epitope may be defined by the presence of high sequence conservation. By "high sequence conservation" is meant a large number of residues are identical in the same molecule expressed in different species (*e.g.*, a hemoglobin molecule from a monkey versus a hemoglobin molecule from a human). Sequence identity is typically measured using sequence analysis software with the default parameters specified therein (*e.g.*, Sequence Analysis Software Package of the Genetics Computer Group, University of Wisconsin Biotechnology Center, 1710 University Avenue, Madison, WI 53705).

Even more preferably, a T cell epitope may be defined by having a lower B-factor value relative to the B-factor values of the entire polypeptide. By "low B-factors" is meant a lower than average disorder or higher than average strong diffraction of X-rays (as determined by X-ray crystallography). Most preferably, a T cell epitope may be defined by having high amide protection. By "high amide protection" is meant an amide protection factor that is higher than 10^4 in a native protein or higher than the average amide protection factor for all residues in a denatured protein.

By "amide protection factor" is meant the rate of hydrogen exchange with solvent deuterons of the amide in an unstructured model peptide divided by the rate of hydrogen exchange for the amide in the context of the protein.

By "unstable polypeptide segment" is meant a chain of amino acids which shows increased flexibility compared to the entire polypeptide, of which the chain of amino acids is a segment. An unstable polypeptide segment is preferentially subject to proteolytic cleavage during antigen processing, and is preferably at least twelve amino acid residues in length. Preferably, no more than 30% of the amino acids of an unstable polypeptide segment are hydrophobic amino acids (*i.e.*, preferably not more

than 30% of the residues are isoleucine, leucine, valine, tyrosine, phenylalanine, tryptophan, threonine, or methionine residues. An unstable polypeptide segment may be defined by a low NMR order parameter (S^2). A low NMR order parameter is an NMR order parameter which is less than 0.8, and is preferably lower than 0.7. By "low NMR order parameter" is meant the low degree of order exhibited by NMR-active nuclei on the time scale of picoseconds to nanoseconds as determined by NMR relaxation processes. Preferably, an unstable polypeptide segment may be defined as having a lower than average hydrophobicity (as determined by a Kyte-Doolittle scale) relative to the average hydrophobicity of the entire polypeptide. More preferably, an unstable polypeptide segment may be defined by the presence of low sequence conservation. By "low sequence conservation" is meant a low number of residues are identical in the same molecule expressed in different species (*e.g.*, a hemoglobin molecule from a monkey versus a hemoglobin molecule from a human). Sequence identity is typically measured as described above.

Even more preferably, an unstable polypeptide segment may be defined by having high average B-factor value relative to the average B-factor value of the entire polypeptide. By "high B-factors" is meant a higher than average disorder or lower than average strong diffraction of X-rays (as determined by X-ray crystallography). Most preferably, an unstable polypeptide segment may be defined by having low amide protection. By "low amide protection" is meant an average amide protection factor that is lower than 10^4 in a native protein or lower than the average amide protection factor for all residues in a denatured protein.

By "antigen" is meant a protein or polypeptide that may be derived from any source.

By "adjuvant" is meant a substance which, when given with an antigen, enhances the immune response to that antigen.

By "inserted by artifice" is meant a protein or polypeptide that has been manipulated by standard techniques for altering polypeptide sequences or nucleic acids encoding polypeptide sequences (including, without limitation, site specific

mutagenesis, deletion, point mutation, homologous recombination, PCR, endonuclease digestion, and ligation), such that the amino acid sequence of the resulting protein has been altered from its naturally-occurring form. Preferably, a protein has been inserted by artifice to include as unstable polypeptide segment. It will be understood that a protein that has been inserted by artifice may not necessarily be greater in length than the naturally occurring protein from which it was derived. For example, a protein may be inserted by artifice, such that a portion of the amino acid sequences of the naturally-occurring protein are deleted or altered.

By "protein" or "polypeptide" is meant any chain of more than two amino acids, regardless of post-translational modification such as glycosylation or phosphorylation.

By "substantially pure polypeptide" (or "substantially pure protein") is meant a polypeptide that has been separated from the components that naturally accompany it. Typically, the polypeptide is substantially pure when it is at least 60%, by weight, free from the proteins and naturally-occurring organic molecules with which it is naturally associated. Preferably, the polypeptide is an antigen or a T cell epitope thereof that is at least 75%, more preferably at least 90%, and most preferably at least 99%, by weight, pure. A substantially pure polypeptide may be obtained, for example, by extraction from a natural source (*e.g.* a fibroblast, neuronal cell, or lymphocyte) by expression of a recombinant nucleic acid encoding a polypeptide, and engineered antigen polypeptide, or T cell epitope thereof, or by chemically synthesizing the polypeptide. Purity can be measured by any appropriate method, *e.g.*, by column chromatography, polyacrylamide gel electrophoresis, or HPLC analysis.

A polypeptide is substantially free of naturally associated components when it is separated from those contaminants which accompany it in its natural state. Thus, a polypeptide which is chemically synthesized or produced in a cellular system different from the cell from which it naturally originates will be substantially free from its naturally associated components. Accordingly, substantially pure

polypeptides not only includes those derived from eukaryotic organisms but also those synthesized in *E. coli* or other prokaryotes.

By "pharmaceutically acceptable carrier" is meant a carrier which is physiologically acceptable to the treated mammal while retaining the therapeutic properties of the compound with which it is administered. One exemplary pharmaceutically acceptable carrier is physiological saline. Other physiologically acceptable carriers and their formulations are known to one skilled in the art and described, for example, in Remington's Pharmaceutical Sciences, (18th edition), ed. A. Gennaro, 1990, Mack Publishing Company, Easton, PA.

By "pathogen" is meant an agent which can cause disease. Exemplary pathogens include viruses (*e.g.*, HIV) and parasites (*e.g.*, helminths, bacteria, and protozoa).

By "allergen" is meant an antigen that generates an allergic reaction.

By "tolerogen" is meant an antigen that inhibits an antigen-specific immune response.

By "immunogen" is meant an antigen that generates an antigen-specific immune response. The greater the immunogenicity of an antigen, the greater the immune response that antigen can elicit.

By "hydrophobic amino acid" is meant an amino acid residue which is aversive to water and forms clusters with like residues. Exemplary hydrophobic amino acids include isoleucine (Ile), leucine (Leu), valine (Val), tyrosine (Tyr), phenylalanine (Phe), Tryptophan (Trp), threonine (Thr), and methionine (Met).

By a "TH1" response is meant a $CD4^+$ helper T lymphocyte response that results in the activation of $CD8^+$ cytotoxic T lymphocytes. TH1 helper T cells produce a number of characteristic cytokines including interleukin-2 (IL-2), interleukin-12 (IL-12), and interferon-γ (IFN-γ).

By a "TH2" response is meant a $CD4^+$ helper T lymphocyte response that results in the activation of antibody-producing B lymphocytes. TH2 helper T cells produce a number of characteristic cytokines including interleukin-4 (IL-4) and

interleukin-10 (IL-10).

Brief Description of the Drawings

Fig. 1A is a schematic diagram of the amino acid sequence of the Hsp10 protein, *M. leprae* cpn10. The T cell epitope at residues 26-42 is shown in solid black (Kim *et al.*, J. Immunol. 159: 335-343, 1997), and the mobile loop at residues 19-35 is shown in speckled black.

Fig. 1B is an X-ray crystal structure of *E. coli* cpn10 (Hunt *et al.*, Nature 379: 37-45, 1996). Grayscale indicates disorder; low B-factors are in dark gray and high B-factors are in light gray. The mobile loop is indicated.

Fig. 2A is a schematic diagram of the amino acid sequence of staphylococcal nuclease. T-cell epitopes are shown in black (Finnegan *et al.*, J. Exp. Med. 164: 897-910, 1986).

Fig. 2B is a series of three panels charting the positions in staphylococcal nuclease of log (protection factor) (Loh *et al.*, Biochemistry 32: 11022-11028, 1993), order parameters (Kay *et al.*, Biochemistry 28: 8972-8979, 1989), and B-factors ($Å^2$) for the amide nitrogen atoms (Brookhaven Protein Data Bank: 1STN).

Fig. 3A is a schematic diagram of the amino acid sequence of hen egg lysozyme. T-cell epitopes are shown in black (Gammon *et al.*, J. Exp. Med. 173: 609-617, 1991).

Fig. 3B is a series of three panels charting the positions in hen egg lysozyme of log (protection factor), order parameters (Buck *et al.*, Biochemistry 34: 4041-4055, 1995), and B-factors ($Å^2$) for the amide nitrogen atoms (Brookhaven Protein Data Bank: 2LYM). Hydrogen exchange protection factors were generated from the data

in Radford *et al.* (Prot. Struct. Funct. Genet. 14: 237-248, 1992) by dividing exchange rates observed at pH 7.5 and 30°C by 10 seconds^{-1}.

Fig. 4 is a series of three schematic diagrams for the amino acid sequences of staphylococcal nuclease, hen egg lysozyme, and cytochrome c, wherein the T-cell epitopes are shown in solid black, and flexible loops are shown in speckled black above the T-cell epitopes.

Fig. 5A is a graph showing the correlation of lysozyme's T cell epitopes with unstable segments at various offsets. The maximum negative correlation is achieved at an offset of -6.

Fig. 5B is a scattergram of lysozyme T cell epitope score versus log (protection factor) for offset equal to -6.

Fig. 6 is a series of schematic diagrams showing the relationship of immunodominant regions to solvent-exposed, proteolytically sensitive sites in HIV gp120.

Fig. 7 is a graph showing the correlation of T-cell epitopes with the change (slope) of average hydrogen exchange (HX) protection factors in hen egg lysozyme. The epitopes were scored by residue as either included (1) or not included (0) in an epitope identified by Gammon *et al.* (J. Exp. Med. 173: 609-617, 1991), and then the positional scores were averaged with a moving window of 11 residues. The HX protection factors from Radford *et al.* (Proteins: Struct. Funct. Genet. 14: 237-248, 1992) were averaged for a 3-residue window, the value of the preceding position was subtracted, and the resulting slopes were averaged with a 9-residue window. Epitopes are found in regions of positive slope of HX protection factors. The Pearson correlation coefficient for the segment 10-106 is 0.50.

Fig. 8 is a graph showing the correlation of T-cell epitopes with adjacent regions having high crystallographic B-factors in lysozyme. The epitopes were defined as in Fig. 7. B-factors for backbone amide nitrogens from the Brookhaven Database (2lym) were averaged with a 5-residue window. Epitopes are in the 15-residue segment from the C-terminal edge of the region of high B-factors. A region of high B-factors is designated as having higher than the average value (14.4, indicated by the horizontal line) for all sequence positions.

Fig. 9 is a graph showing the correlation of T-cell epitopes with sequence conservation in lysozyme. Epitopes were defined as in Fig. 7. Sequence conservation was evaluated for seventy lysozyme sequences. The most popular residue at each position was determined, and then scored for its use at that position in each sequence. Conservative substitutions were counted the same as identical matches. Conservation scores were averaged with a 9-residue window. Epitopes are located in regions with above average conservation (>0.69). Although the correlation does not appear to be very striking, when the epitope score is correlated with conservation 8 residues toward the N-terminus, the Pearson correlation coefficient is -0.57. Thus, epitopes are correlated with poorly conserved regions (processing sites) 8 residues toward the N-terminus.

Fig. 10 is a graph showing the correlation of T-cell epitopes with hydropathy in lysozyme. Epitopes were defined as in Fig. 7. Epitopes tended to occur in regions with above average hydropathy values (11-residue window).

Fig. 11A is a graph showing the correlation of T-cell epitopes with sequence conservation in HIV gp120. Epitope identifications were obtained from the HIV Immunology Database (http://hiv-web.lanl.gov), which includes a compilation of epitopes published prior to the year 1995 (see Fig. 11B). Sequence positions found in multiple epitopes are scored multiple times. Epitope scores were smoothed by an 11-

residue averaging window. Sequence conservation was evaluated for twenty-three gp120 sequences. The most popular residue at each position was determined, and then scored for its use at that position in each sequence. Conservative substitutions were counted the same as identical matches. Conservation scores were averaged with a 11-residue window. Epitopes are located in regions with above average conservation (>0.87).

Fig. 11B is the sequence of HIV gp120 (SEQ ID NO: 1) obtained from the HIV Immunology Database (http://hiv-web.lanl.gov) showing the 36 epitope identifications of epitopes published prior to the year 1995.

Fig. 12 is a graph showing the correlation of T-cell epitopes with hydropathy in HIV gp120. Epitope scores were evaluated as described in Fig. 11. The hydropathy profile was calculated using the Kyte and Doolittle scale with a window size of 11. Little or no correlation was evident.

Figs. 13A, 13B, and 13C are graphs showing the combination of predictive methods based on MHC binding and preferred processing at poorly conserved regions. In Fig. 13A, the raw experimental epitope scores are shown. In Fig. 13B epitopes were predicted by the EpiMatrix analysis available at http://hiv-web.lanl.gov/immuno/articles/LANL.html (no geographic bias in MHC preferences, 20% match to motifs allowed). In Fig. 13C, the EpiMatrix scores were set to zero when the sequence conservation at that residue fell below the average for the whole protein. Note in particular the elimination of predicted epitopes in the segment 130-200.

Fig. 14 is a graph showing the correlation of T-cell epitopes with the combined epitope prediction. Experimental epitope scores are as defined in Fig. 7, and predicted epitope scores from Fig. 13C were smoothed by an averaging window

of 11 residues. Although the Pearson correlation coefficient is below 0.2, visual inspection showed a very good qualitative correlation. The poor quantitative correlation may be due to the unsystematic identification of T-cell epitopes. Some epitopes may be over-represented in the sample for reasons unrelated to processing or presentation.

Fig. 15 is a series of schematic diagrams of the amino acid sequence of HIV gp120 indicating (with arrows) the sites of insertion for an unstable polypeptide segment, the human Hsp10 mobile loop.

Figs. 16A-16D depict the immunodominant helper T cell epitopes in bacteriophage T4 Gp31. Fig. 16A is a bar graph depicting the Stimulation Index (SI) of 15-mer synthetic peptides derived from Gp31. Fig. 16B is a graph showing the hydropathy values of Gp31. Shown below the graphed values is the amino acid sequence of Gp31 (SEQ ID NO: 2). Fig. 16C is a schematic of the 15-mer peptides derived from Gp31. The immunodominant peptides (*i.e.*, the peptides having a SI value greater than 1) are shown in black. Fig. 16D is a graph showing the B-factor values for Gp31.

Detailed Description of the Preferred Embodiment

T cell epitopes are peptides generated by the processing of antigens that are presented by MHC proteins to T cells. The spectrum of immunogenic epitopes from a given antigen is limited by the range of peptides generated by processing and by the binding preferences of the MHC proteins. However, many more peptides of a given antigen are predicted to bind to MHC proteins than are actually observed to restimulate T cells that were primed with the naturally processed antigen. Thus, antigen processing is biased toward production of certain peptides. I have determined that preferential proteolytic cleavage in poorly ordered regions of a substrate polypeptide provides a mechanism for biased production of certain peptides. I have

developed criteria to detect and design immunodominant T cell epitopes.

Immunodominant helper T-cell epitopes near highly flexible loops

Mobility of protein regions may be demonstrated by NMR and protease sensitivity (see, for example, determination for the GroES loop in Landry *et al.*, Nature 364: 255-258, 1993). In an attempt to understand and thus predict patterns of immunodominance, I examined features of protein dynamics for their correlation with T-cell epitopes. My finding that T-cell epitopes coincide with highly flexible polypeptide loops in the E2 subunit of the mitochondrial pyruvate dehydrogenase complex (PDC), as well as in *Mycobacterium leprae* chaperonin 10 (cpn10) suggested that flexible segments could be targets for processing and presentation. The lipoyl domains of the PDC E2 subunit contain the immunodominant epitopes known to be associated with the autoimmune disease primary biliary cirrhosis. The lipoyl domains are sufficiently flexible to be observed by solution-phase nuclear magnetic resonance (NMR) despite the large size of the PDC, and they are preferentially cleaved by proteases (Perham and Duckworth, Nature 292: 474-477, 1981).

The *M. leprae* cpn10 protein, an Hsp10 protein, contains one immunodominant epitope (Kim *et al.*, J. Immunol. 159: 335-343, 1997) (Fig. 1A) which overlaps a mobile loop which were identified in the homologous sequence of the *Escherichia coli* cpn10, GroES (Fig. 1B). No electron density corresponding to the mobile loop was observed in the *M. leprae* cpn10 crystal structure, presumably due to disorder, and density was observed for the loop of only one of seven subunits in the crystal structure of *E. coli* GroES (Hunt *et al.*, Nature 379: 37-45, 1996).

Structural instability mapped by crystallographic B-factors

To obtain further support for my hypothesis that flexible segments direct antigen processing, results of T-cell fine-mapping studies were compared with dynamic properties of staphylococcal nuclease and hen egg lysozyme, two proteins whose immunology, structure, and dynamics are well studied. The dynamic

properties consisted of B-factors, which are assigned to each atom during X-ray crystallographic studies on a molecule. Poorly ordered atoms diffract X-rays weakly at high resolution; because of this effect, the degree of disorder can be estimated in the course of the quantitative refinement of the crystal structure, which seeks to optimize the agreement between the observed diffraction intensities and the molecular model (Ringe and Petsko, Methods Enzymol. 131: 389-433, 1986). Crystallographic B-factors (also known as temperature factors) weight the contribution of atoms in the back-calculation of electron density from the model structure. High B-factors correspond to weak density. Information concerning the origin of high B-factors is not directly available from the data. High B-factors can result from time-averaged fluctuations or static inhomogeneity.

T cells primed with native staphylococcal nuclease gave the strongest proliferative responses to epitopes adjacent to a segment with highly elevated B-factors (Fig. 2A and bottom panel of Fig. 2B). In lysozyme, seven determinant cores have been identified, and there is a similar number of peaks in the B-factor profile (Fig. 3A and bottom panel of Fig. 3B). In general, epitopes occurred in well-ordered regions between flexible sites, consistent with a processing mechanism in which proteolytic nicking of the native protein generates fragments that ultimately are presented in the MHC proteins. However, elevated B-factors also occurred at sites where T-cell epitopes have not been discovered (*e.g.*, residues 25-32 in staphylococcal nuclease), and epitopes are found where B-factors are not elevated (*e.g.*, epitope 81-100 in staphylococcal nuclease). Aside from the fact that proteolytic processing may not be the only determinant of immunodominance, the occasional failure of B-factors to predict epitopes suggested that elevated B-factors are not the best predictor of proteolytically sensitive sites. Indeed, analysis of B-factors failed to predict a proteolytically sensitive site in trypsinogen (Novotny and Bruccoleri, FEBS Lett. 211: 185-189, 1987).

Structural instability mapped by NMR relaxation parameters

Nuclear magnetic resonance (NMR) has emerged as a powerful technique for the measurement of dynamic properties of proteins in solution. NMR often is conceptualized as a revolving charge, or spin. Nuclear spins aligned with a magnetic field can be excited by an electromagnetic pulse that reorients the spins perpendicular to the magnetic field. Realignment of the spins with the magnetic field occurs by two mechanisms of relaxation: longitudinal (spin-lattice) relaxation and transverse (spin-spin) relaxation. When the molecular motion is isotropic (rotation equally fast in all directions), the speed of the motion affects longitudinal and transverse relaxation rates in a very predictable way; thus, the speed of molecular motion can be estimated from NMR relaxation rates.

In the context of NMR, the nuclear Overhauser effect (NOE) refers to the transfer of magnetization between two nuclei, a form of longitudinal (spin-spin) relaxation. The magnitude of the NOE is affected by the distance between the nuclei and the rate of molecular motion. ^{1}H-^{1}H NOEs are used determine the structure of protein molecules. As referred to herein, amide ^{15}N-^{1}H NOEs are used to measure motion. The distance dependence is not a factor because the ^{15}N and ^{1}H atoms are bonded to each other and therefore nearly equidistant for all amide groups.

A simple and powerful set of NMR experiments are gaining wide popularity for studying protein dynamics. According to this strategy, the protein is uniformly labeled with [^{15}N]nitrogen, which is an NMR-active nucleus, and all of the experiments are "isotope-edited", in the sense that only nitrogen-bound hydrogen atoms are observed. This serves to simplify the spectra, and focuses attention on the dynamic properties of the backbone amide nitrogens. In this way, the technique provides a probe of molecular dynamics at each amino acid residue in the protein, except proline since this residue lacks the amide proton.

Typically, three NMR parameters are measured: the transverse relaxation rate, the longitudinal relaxation rate, and the ^{15}N-^{1}H nuclear Overhauser effect. Then, by combining these parameters in equations that describe the NMR relaxation processes, one can estimate the extent of conformational fluctuations that occur on time-scales

ranging 10^{-12} - 10^{-3} seconds. Since a physical description of the conformational fluctuations involves more variables than can be measured, the extent of disorder is often analyzed by the "model-free" approach, which yields the square of the order parameter, S^2 (see Lipari and Szabo, J. Amer. Chem. Soc. 104: 4546-4559, 1982; Lipari and Szabo, J. Amer. Chem. Soc. 104: 4559-4570, 1982).

An order parameter generally describes the correlation of two vectors in space. An order parameter derived from protein ^{15}N-^{1}H relaxation data describes the correlation of a vector along the NH bond with a fixed axis of the molecule. A low order parameter indicates motion that is independent of the molecular rotation. An S^2 value of unity corresponds to atomic motion that is perfectly correlated with the overall molecular rotation. Typical values for well-ordered regions of a protein are in excess of 0.8.

Several T-cell epitopes in staphylococcal nuclease and lysozyme lie adjacent to highly mobile segments. Residues 45-49, 69-74 and 115-120 in lysozyme exhibit S^2 values below 0.8, and these segments all have high B-factors in the crystal structure (Fig. 2B, middle panel). Although its order parameters could not be analyzed, the segment 42-49 in staphylococcal nuclease is highly mobile by NMR (Torchia *et al.*, Structure 28: 5509-5524, 1989), and this is part of a segment with high B-factors. Resonances for this segment were difficult to detect and assign, probably because conformational fluctuations occur close to the NMR time-scale (approximately 10^{-3} seconds) (Kay *et al.*, Biochem. 28: 8972-8979, 1989). Thus, the identification of disordered segments using order parameters revealed only a subset of the segments identified by elevated B-factors. Elevated B-factors may sometimes indicate disorder with fluctuations of longer time-scale than 10^{-3} seconds. Nevertheless, some epitopes were predicted neither by order parameters nor B-factors. Perhaps these epitopes were also directed by preferred proteolytic nick sites, but the conformational transitions that favor nicking are very slow, and the population of alternate conformations is so small that B-factors are not significantly affected.

Structural instability mapped by hydrogen-deuterium exchange NMR

Disorder characterized by long time-scale structural fluctuations can be detected by NMR-based measurement of hydrogen-deuterium exchange (HX) for backbone amide groups. With the additional sensitivity of this technique, which is described below in detail, a striking correlation was found between regions of structural stability and T-cell epitopes (Figs. 2B and 3B, top panels), suggesting that the intervening regions of instability defined the ends of T-cell epitopes. The HX technique effectively analyzes structural stability, in that a given amide group is protected against exchange while in a hydrogen-bonded structure, but readily undergoes exchange when the segment transiently unfolds. Amide groups of fully unfolded polypeptides exchange in approximately 0.1 seconds at room temperature and neutral pH. Typically, backbone amide HX protection is analyzed by two-dimensional NMR spectra. Amide proton occupancy is monitored with a series of experiments over time after the protein is dissolved in deuterium oxide. During this period, amide protons exchange for the NMR-inactive deuterons. Generally, amide protons involved in hydrogen bonds exhibit significant HX protection; thus, the signal persists for at least a few minutes, which is the length of time necessary to acquire the first spectrum. The most stable segments of proteins are highly protected against HX, with signals persisting for months. The time-scale of structural transitions consistent with poor HX protection (minutes to hours) is comparable with the time-scale of intracellular processing events leading to peptide presentation; thus, it can be expected that preferred proteolytic sites will coincide with regions of poor HX protection. Since HX is sensitive to fluctuations on relatively long time-scales, segments with faster conformational motions, which may be identified by order parameters or B-factors, should be a subset of segments identified by poor HX protection. Indeed, this was the case for staphylococcal nuclease and lysozyme. Some regions of poor HX protection emerged where B-factors and order parameters were inconclusive; for example, residues 76-88 in staphylococcal nuclease and 86-91 in lysozyme.

For three well characterized proteins of similar size, hen egg lysozyme, staphylococcal nuclease, and cytochrome c, the position of T-cell epitopes have been compared with the positions of locally unstable segments (Fig. 4). For lysozyme and staphlococcal nuclease, a segment was scored as flexible if the amide protection factors were 10^3 or less for a stretch of six or more residues. (The protection factor is the hydrogen-deuterium exchange rate for an amide group in an unstructured peptide divided by the exchange rate of a specific amide group in the protein.) For cytochrome c, assignment of stable and unstable segments was as specified on the basis of amide protection factors (Bai *et al.*, Science 269: 192-197, 1995). For lysozyme, T-cell epitopes were determined by proliferation assays using three series of peptides that were 10, 12, and 15 residues in length and spanning the lysozyme sequence in one-residue steps (Gammon *et al.*, J. Exp. Med 173: 609-617, 1991). Maps were determined for B10.A, BALB.B, and BALB/c strains of mice. Only epitopes determined for B10.A and BALB/c strains were considered in this analysis since all epitopes determined for the BALB.B strain overlapped those determined with the B10.A strain. It is likely that detailed differences between the B10.A and BALB.B epitopes derive from differences in MHC alleles. For staphylococcal nuclease, the peptide series was composed of 20-mers having 10-residue overlaps (Finnegan *et al.*, J. Exp. Med 164: 897-910, 1986). For staphylococcal nuclease, additional peptides spanning 61-80 and 91-110 were stimulatory but were not included in the present analysis because they overlapped with 51-70 and 81-100, respectively. These slightly different but overlapping epitopes may be explained by MHC allele specificity. For cytochrome c, only a few non-overlapping 15-mers were used to produce the T-cell epitope map (Mamula *et al.*, J. Immunol. 149: 789-795, 1992).

In most cases, the epitopes overlapped the intervening unstable regions and were displaced to the C-terminal side. This pattern was consistent with a mechanism of antigen processing that is initiated via proteolytic cleavage near the center of a locally disordered polypeptide segment, followed by binding of the proximal

sequence of the C-terminal product in the MHC class II cleft. Presumably, fine-tuning of the binding site within a particular peptide would occur at the level of MHC recognition. Further proteolytic trimming of the peptide might occur in the context of the peptide-MHC class II complex.

The number and position of T-cell epitopes correlated with the number and position of unstable segments identified by amide protection factors. The apparent correlation between unstable segments and T-cell epitopes is borne out by statistical analysis of the data for lysozyme (Figs. 5A and 5B) and staphylococcal nuclease. Each residue was assigned a T-cell epitope "score" (1 or 0) on the basis of whether or not it was part of an epitope. Pearson correlation coefficients (r) were calculated for epitope scores and log (protection factors). Correlation coefficients also were calculated after smoothing the data with a moving average using window sizes ranging from 3 to 11. Without offsetting the epitope scores, the correlation was near zero for both lysozyme and staphylococcal nuclease. However, for lysozyme, the correlation rose to 0.47 (3-residue window) when the epitope score was offset toward the C-terminus, and the correlation increased to an even greater negative magnitude of -0.56 (11 residue window) when the epitope score was offset toward the N-terminus (Fig. 5A). On average, lysozyme's T-cell epitopes lie 6 residues downstream from unstable segments and 4 residues upstream from stable segments (center-to-center). For staphylococcal nuclease, a maximum positive correlation ($r = 0.5$) was obtained with an offset of +10, and maximum negative correlation ($r = -0.43$) with an offset of -11. The slightly larger offsets obtained for staphylococcal nuclease may be related to the fact that fewer and larger peptides were used in the mapping study.

The observed correlation is consistent with a mechanism of antigen processing in which proteolytic cleavage occurs near the center of a locally disordered polypeptide segment followed by binding of the N-terminal sequence of the C-terminal product in the MHC II protein cleft. Subsequently, further proteolytic processing trims the C-terminal end to the final size, wherein the portion of the peptide bound by MHC is protected against further degradation. Hence, MHC

peptide binding occurs before antigen processing is complete. Thus, I deduced that peptides adjacent to the most readily cleaved sites in proteins should be preferentially presented.

The results described above suggest that T-cell epitopes occur in the regions immediately C-terminal from unstable segments; and within a region, exact sequences of the epitopes can be predicted by the selectivity of MHC alleles. The strength of the correlation between flexible sites and epitopes can be assessed by the coefficient of determination (r^2). The correlation coefficient of -0.56 obtained for lysozyme (offset = -6) translates to a r^2 of 0.31, which means that 31% of the variance in T-cell epitope scores is explained by the correlation with amide protection factors. The remaining 69% of the variance must be explained by other factors, including selectivity at the level of peptide binding to MHC proteins.

Helper T-cell Epitopes in gp120

As was observed for lysozyme, *M. leprae* cpn10, and staphylococcal nuclease, helper T-cell epitopes in *H*uman *I*mmunodeficiency *V*irus (HIV) gp120 tend to cluster near sites that may be preferentially cleaved during antigen processing (Fig. 6). Epitopes were defined using a variety of T-cell stimulation systems, for example, with lymphocytes from draining lymph nodes of mice immunized with native gp120 (Cease *et al.*, Proc. Natl. Acad. Sci. USA 84: 4249-4253, 1987), peripheral blood lymphocytes from humans immunized with vaccinia virus expressing gp120 (Berzofsky *et al.*, Nature 334: 706-708, 1988), and peripheral blood lymphocytes from HIV patients (Clerici *et al.*, Nature 339: 383-385, 1989). Data have been collected and published on the World Wide Web in the HIV Immunology Database, Los Alamos National Labs (http://hiv-web.lanl.gov). Epitopes are broadly distributed over the C-terminal half of gp120. Fewer epitopes occur in the N-terminal half of the protein, although there is a cluster in the region, 101-119, including the "T2" epitope (Cease *et al.*, *supra*). Overlapping epitopes may be grouped into eight regions (shown as gray boxes in Fig. 6) that encompass most of the gp120 sequence: 31-54, 64-84,

101-119, 203-269, 273-301, 306-369, 417-453, and 457-502. However, several much shorter segments are over-represented in the sample of reported epitopes. We define immunodominant sequences as those occurring in at least four epitopes: 104-115, 225-236, 294-297, 311-349, 426-440, and 486-500. The number of immunodominant regions (shown as black boxes in Fig. 6) is very similar to the number of solvent-exposed segments (shown as white boxes in Fig. 6) identified by monoclonal antibodies that bind to linear epitopes in native gp120 (Moore *et al.*, J. Virol. 68: 469-484, 1994), and the immunodominant regions tend to be adjacent to the solvent-exposed segments. Three of these solvent-exposed segments also were preferentially susceptible to proteolysis. Taken together, these observations demonstrated that proteolytic nicking targets presentation of nearby sequences in HIV gp120.

Implications

My discovery that local instability within a protein directs immunodominance provides a framework for understanding the interaction of an antigen with the cellular processing mechanism. Local instability enhances susceptibility to proteolytic attack and thus, I believe, provides a mechanism for preferential presentation of adjacent peptides. Such a mechanism allows the design of antigens with enhanced immunogenicity, and also the detection of immunodominant T cell epitopes in a group of antigens. These methods may allow the treatment and prevention of diseases caused by pathogens. These methods also allow the detection of auto-reactive T lymphocytes in patients suffering from autoimmune diseases. In addition, by the ability to detect and/or design immunodominant epitopes, tolerogens may be designed to prevent or alleviate allergy or autoimmune diseases.

In vivo, pathogens might take advantage of bias in the processing mechanism to suppress the development or activation of deleterious T cells. For example, one can imagine a T-cell epitope that is preferentially presented in B cells but is consistently bypassed in macrophages. This strategy as described herein may be

particularly effective for parasites that bear highly repeated, structurally disorganized immunodominant surface antigens. My detection of the correlation of instability with T cell epitopes suggests that new T cell epitopes could be introduced, or the immunogenicity of existing epitopes could be enhanced in an antigen, by incorporating unstable segments near the desired target epitopes.

Criteria for prediction of T-cell epitopes

The data provided herein indicate that T-cell epitopes tend to be most abundant on the C-terminal flank of poorly ordered antigen segments. An absence of epitopes within poorly ordered segments is consistent with their preferential proteolysis. The apparent bias toward presentation of the C-terminal product may be due to an inherent preference by the MHC proteins or other intrinsic feature of the presentation mechanism.

T-cell epitopes (also referred to herein as "epitopes") can be predicted in the following ways (in order of most effective to least effective):

I. Hydrogen-deuterium exchange NMR

When a protein is dissolved in deuterium oxide, the amide protons exchange with deuterons at a rate which is limited by the degree of local ordered structure. Typically, amide exchange data is reported as hydrogen exchange (HX) protection factors. An HX protection factor is defined as the rate of exchange of the amide in a disordered model peptide divided by the rate observed in the context of the protein. Regions of low amide protection are preferred processing sites. Epitopes lie in the adjacent sequences. Several criteria have been developed for identification of T-cell epitopes on the basis of HX protection. The utility of each criterion may depend on the completeness of the HX protection data and the conditions under which the protection data is obtained. All averages of HX protection factors are evaluated using the logarithms of the HX protection factors.

a. Epitopes are located in regions where the average HX protection factor increases over a length of the polypeptide chain, in the amino-terminal to carboxy-

terminal direction of the polypeptide. The average change of average HX protection factor is evaluated as follows: (i) the HX protection factors are "smoothed" by assigning to each position the average of the HX protection factor for that position and the two flanking positions (a three-residue window); (ii) the change in average value for each position is evaluated by subtracting the value at the previous position; (iii) the change in HX protection is averaged for a window of seven (7) residues. Epitopes are found where the average change in HX protection is positive (Fig. 7). The precise location of the epitope is determined by compatibility of the sequence with binding to MHC proteins.

b. Epitopes are adjacent to processing sites identified by at least six (6) consecutive residues with HX protection factors below 10^4 for the antigen in its native conformational state. Typically, epitopes are within fifteen (15) residues of the C-terminal edge of the region of low protection.

c. Epitopes are located between regions of low HX protection or within regions of high HX protection. High and low HX protection are defined relative to the average protection factor for all amide groups in the protein. This criterion replaces that of Ib for amide exchange data obtained for the antigen in a denatured state.

II. X-ray crystallographic B-factors

B-factors are assigned to each atom during the refinement phase of the determination of molecular structures by X-ray crystallography. Poorly ordered atoms diffract x-rays weakly. The B-factor adjusts the contribution of the atom to the electron density map, with a higher B-factor yielding a lower density. Poorly ordered segments have high average B-factors for amide nitrogen atoms. Regions with high B-factors are preferred processing sites. The average B-factor is evaluated with an averaging window with an optimal length of five (5) residues. Epitopes are adjacent to and on the C-terminal side of antigen segments characterized by high average amide nitrogen B-factors (Fig. 8). Typically, epitopes are within fifteen (15) residues of the C-terminal edge of the region of high B-factors.

III. Sequence conservation

Residues that are important for specifying the three-dimensional structure of a protein tend to be conserved as identical among homologs in different species or viral strains. The sequence is less constrained where the structure is poorly ordered, and fewer residues are conserved as identical. Therefore, regions of low sequence conservation are preferred processing sites. Epitopes are located between regions of low sequence conservation or within regions of high sequence conservation (Figs. 9, 11, 13A, 13B, 13C, and 14). Conservation at a given sequence position is evaluated as the number of sequences bearing the most common residue divided by the total number of sequences. Conserved regions have a higher average conservation than the average for all sequence positions in the protein. Regional conservation is analyzed by an averaging window seven to fifteen (7-15) residues in length.

IV. Hydropathy

Hydrophobic amino acids tend to participate in formation of a protein's hydrophobic core, which is a well-ordered portion of its three-dimensional structure. Hydrophilic segments are less well ordered and are preferred processing sites. Antigen segments involved in the core structure have higher hydrophobicity, as analyzed by an averaging window seven to fifteen (7-15) residues in length using the Kyte-Doolittle scale (Kyte and Doolittle, J. Mol. Biol. 157: 105-132, 1982). Epitopes are located between regions of low hydrophobicity or within regions of high hydrophobicity, defined relative to the average hydrophobicity for all residues in the protein (Figs. 10 and 12).

V. NMR order parameters, accessibility to antibodies, proteolytic nicking

NMR order parameters, evaluated by measurement of NMR relaxation processes, define the degree of motion experienced by NMR-active nuclei on the time scale of picoseconds to nanoseconds. Regions of a protein that exhibit low order parameters are preferred processing sites, and regions of high order parameters contain epitopes. Some antibodies raised against protein antigens recognize both the antigen and a synthetic peptide corresponding to a sequence of the antigen. It is likely

that these antibodies recognize the same sequence in both molecules. If these antibodies react with the antigen in the native state, the sequence must be exposed; whereas if the antibody reacts only with the denatured antigen, the sequence must be sequestered. Exposed sequences are preferred processing sites, and sequestered sequences contain epitopes. Proteolytic nick sites can be defined for an antigen by peptide mapping (isolation of the proteolytic products followed by sequencing). Proteolytic nick sites are preferred processing sites, and sequences between nick sites contain epitopes.

VI. Additional Criteria

Additional circumstances, such as the particular cell type or the presence of inflammation, may modify the spectrum of peptides generated. For example, local protease sensitivity may be significantly affected by the presence of components that bind to the antigen, such as antibodies or molecular chaperones. Indeed, antigen presentation was found to differentially affected for complexes of antigen with various monoclonal antibodies (Manca *et al.*, J. Immunol. 9: 2893-2898, 1988).

In addition, B-cell receptors might inhibit B-cell antigen processing by blocking the preferred processing sites, or conversely, enhance processing by locally destabilizing the structure. Investigators studying mouse models of autoimmune diseases, experimental allergic encephalomyelitis (Lehmann *et al.*, Nature 358: 155-157, 1992) and insulin-dependent diabetes (Kaufman *et al.*, Nature 366: 69-72, 1993), have observed changes in the pattern of immunodominance over time after the initial immunization, a phenomenon known as "epitope spreading." Migration of the dominant epitopes is thought to arise from successive rounds of antigen processing and presentation by evolving B cell populations (Mamula, M. J., Immunol. Rev. 144: 301-315, 1995; Holmdahl *et al.*, Immunol. Rev. 144: 109-135, 1995). Hence, the evolution of B-cell receptor specificity and antibody affinity maturation could be additional mechanisms that drive changes in T-cell epitope immunodominance.

Development of Vaccines with Engineered T-cell Epitopes

The present invention allows the generation of improved vaccines by insertion of unstable polypeptide segments into an antigen (*e.g.*, HIV gp120). The inserted segments are designed to enhance proteolytic processing at the site of insertion, resulting in enhanced presentation of epitopes from the antigen. Since presentation seems to be biased toward presentation of the sequence on the C-terminal side of an unstable segment, the insertion may be used to target presentation of a specific antigen sequence, such as one that is highly conserved and thus more likely to be broadly protective against pathogenic variants.

Unstable polypeptide segments are expected to be preferred sites of proteolytic attack during antigen processing into peptides that will be presented in class II MHC antigen presenting proteins (and possibly also in class I antigen presenting proteins). The engineered polypeptide insert should not be shorter than twelve (12) residues or approximately the size of a protease recognition site; the insert should be composed of no more than 30% hydrophobic amino acids (Ile, Leu, Val, Tyr, Phe, Trp, Thr, Met); the insert may contain a sequence recognized by a particular protease (*e.g.*, a sequence recognized by the HIV protease); and the insert should not be predicted to form an ordered structure.

The enhancement of epitope immunogenicity by insertion of unstable segments is expected to be most effective for the priming of $CD4^+$ helper T cells by antigen presenting cells, with presentation restricted by class II MHC proteins. This is the typical mode of presentation of "exogenous" antigens, *e.g.*, subunit vaccines and proteins of bacteria, cell debris, and killed viruses. However, exogenous antigens can be presented in class I MHC proteins, and immunodominant helper T-cell epitopes often coincide with immunodominant cytolytic ($CD8^+$) T-cell epitopes. Thus, enhancement of processing by insertion of unstable segments is a strategy to enhance cytolytic T cell responses to the targeted epitope.

Identification of an Immunodominant Epitope in an Antigen

Unstable polypeptide segments mark sites of enhanced proteolytic processing. In a set of antigen proteins that are otherwise equally available to immune surveillance, the protein with the most poorly ordered polypeptide segment will be the most immunogenic. Furthermore, in a set of antigens derived from a pathogen, the most immunogenic antigen may be the most suitable for development as a subunit vaccine. In a set of self antigens, the most immunogenic protein may contribute to autoimmune dysfunction and symptomology; and thus the epitope would be a candidate for development in a tolerogenic formulation.

Stimulation of T cell proliferation indicates that a previous infection, immunization, or autoimmune process primed a response to the targeted epitope. Proliferation in response to an epitope of a pathogen protein by T cells from an infected individual indicates that the T cells were primed by the natural antigen. Observation of the proliferative response can provide useful information: The proliferative response to the epitope of a pathogen protein confirms that the T cell donor was exposed to the antigen and thus may have been or currently is infected with the corresponding pathogen.

The response to the epitope of a pathogen protein indicates that the epitope is a candidate for immune enhancement in an engineered vaccine. Similarly, a response to the epitope by T cells from an individual immunized with a pathogen protein or inactivated pathogen indicates priming against the epitope; and thus the epitope is a candidate for enhancement in an engineered vaccine. A response to an epitope from a self antigen indicates priming of an autoimmune response. An autoimmune response may be diagnostic of an autoimmune disorder, and the response to the epitope is the basis for design of a tolerogenic treatment with a formulation of the epitope.

For example, the most immunodominant epitopes of the HIV antigen, gp120, have been predicted (see Fig. 6, black boxes). The $CD4^+$ T lymphocytes of an individual who either is, or is suspected to be, infected with HIV may be isolated and used in a proliferation assay to determine responsiveness to the epitope presented on

autologous antigen presenting cells. The presence of T lymphocytes which respond to the immunodominant epitope indicates that individual is infected with HIV. Furthermore, this information allows the assessment of the type of response the HIV infected individual is likely to generate against the virus. For example, once the HIV gp120 responsive $CD4^+$ T lymphocytes are identified, they may be assessed for an ability to stimulate either a cytotoxic T cell-mediated or antibody-mediated immune response through differential cytokine release. Knowledge of infection and type of response generated is useful for the treatment of infection with HIV, and other pathogens.

Generation of a T cell Epitope Alone, or Within an Engineered Recombinant Protein

Once an immunodominant epitope is identified, or is engineered to be immunodominant in an antigen of choice, a variety of methods may be employed to produce the epitope, or epitope-containing antigen. For example, DNA sequences encoding the T cell epitope, or the epitope-containing antigen are introduced into a plasmid or other vector which is then used to transform living cells. Constructs in which the epitope, or epitope-containing antigen DNAs containing the entire open reading frames inserted in the correct orientation into an expression plasmid may be used for protein expression.

Typical expression vectors contain promoters that direct the synthesis of large amounts of mRNA corresponding to the inserted epitope, or epitope-containing antigen nucleic acid in the plasmid bearing cells. They may also include eukaryotic or prokaryotic origin of replication sequences allowing for their autonomous replication within the host organism, sequences that encode genetic traits that allow vector-containing cells to be selected for in the presence of otherwise toxic drugs, and sequences that increase the efficiency with which the synthesized mRNA is translated. Stable long-term vectors may be maintained as freely replicating entities by using regulatory elements of, for example, viruses (*e.g.*, the OriP sequences from the Epstein Barr Virus genome). Cell lines may also be produced which have integrated

the vector into the genomic DNA, and in this manner the gene product is produced on a continuous basis.

Expression of foreign sequences in bacteria such as *Escherichia coli* requires the insertion of the T cell epitope, or epitope-containing antigen nucleic acid sequence into a bacterial or eukaryotic expression vector. This plasmid vector contains several elements required for the propagation of the plasmid in bacteria or eukaryotic cells, and expression of inserted DNA of the plasmid by the plasmid-carrying bacteria or eukaryotic cell. Propagation of only plasmid-bearing bacteria is achieved by introducing in the plasmid selectable marker-encoding sequences that allow plasmid-bearing bacteria to grow in the presence of otherwise toxic drugs. The plasmid also bears a transcriptional promoter capable of producing large amounts of mRNA from the cloned gene. Such promoters may or may not be inducible promoters which initiate transcription upon induction. The plasmid also preferably contains a polylinker to simplify insertion of the gene in the correct orientation within the vector. In a simple *E. coli* expression vector utilizing the lac promoter, the expression vector plasmid contains a fragment of the *E. coli* chromosome containing the lac promoter and the neighboring lacZ gene. In the presence of the lactose analog IPTG, RNA polymerase normally transcribes the lacZ gene producing lacZ mRNA which is translated into the encoded protein, β-galactosidase. The lacZ gene can be cut out of the expression vector with restriction endonucleases and replaced by a T cell epitope, or epitope-containing antigen DNA sequence. When this resulting plasmid is transfected into *E. coli*, addition of IPTG and subsequent transcription from the lac promoter produces epitope, or epitope-containing antigen mRNA, which is translated into epitope, or epitope-containing antigen polypeptides.

Once the appropriate expression vectors containing a epitope, or epitope-containing antigen DNA are constructed, they are introduced into an appropriate host cell by transformation techniques including calcium phosphate transfection, DEAE-dextran transfection, electroporation, micro-injection, protoplast fusion and liposome-mediated transfection. The host cell which are transfected with the vectors

of this invention may be selected from the group consisting of *E. coli*, *pseudomonas, Bacillus subtilus*, or other bacilli, other bacteria, yeast, fungi, insect (using, for example, baculoviral vectors for expression), mouse or other animal or human tissue cells. Mammalian cells can also be used to express the epitope, or epitope-containing antigen using a vaccinia virus expression system described in Ausubel *et al.* (Current Protocols in Molecular Biology, John Wiley & Sons, New York, NY, 1994).

In vitro expression of epitope, or epitope-containing antigen encoded by cloned DNA is also possible using the T7 late-promoter expression system. This system depends on the regulated expression of T7 RNA polymerase which is an enzyme encoded in the DNA of bacteriophage T7. The T7 RNA polymerase transcribes DNA beginning within a specific 23-bp promoter sequence called the T7 late promoter. Copies of the T7 late promoter are located at several sites on the T7 genome, but none is present in *E. coli* chromosomal DNA. As a result, in T7 infected cells, T7 RNA polymerase catalyzes transcription of viral genes but not of *E. coli* genes. In this expression system recombinant *E. coli* cells are first engineered to carry the gene encoding T7 RNA polymerase next to the lac promoter. In the presence of IPTG, these cells transcribe the T7 polymerase gene at a high rate and synthesize abundant amounts of T7 RNA polymerase. These cells are then transformed with plasmid vectors that carry a copy of the T7 late promoter protein. When IPTG is added to the culture medium containing these transformed *E. coli* cells, large amounts of T7 RNA polymerase are produced. The polymerase then binds to the T7 late promoter on the plasmid expression vectors, catalyzing transcription of the inserted cDNA at a high rate. Since each *E. coli* cell contains many copies of the expression vector, large amounts of mRNA corresponding to the cloned cDNA can be produced in this system and the resulting protein can be radioactively labelled. Plasmid vectors containing late promoters and the corresponding RNA polymerases from related bacteriophages such as T3, T5, and SP6 may also be used for *in vitro* production of proteins from cloned DNA. *E. coli* can also be used for expression by infection with M13 Phage mGPI-2. *E. coli* vectors can also be used with phage lambda regulatory

sequences, by fusion protein vectors, by maltose-binding protein fusions, and by glutathione-S-transferase fusion proteins.

Eukaryotic expression systems permit appropriate post-translational modifications to expressed proteins. Transient transfection of a eukaryotic expression plasmid allows the transient production of an epitope, or epitope-containing antigen by a transfected host cell. A T cell epitope, or epitope-containing antigen may also be produced by a stably-transfected mammalian cell line. A number of vectors suitable for stable transfection of mammalian cells are available to the public (*e.g.*, see Pouwels *et al.*, Cloning Vectors: A Laboratory Manual, 1985, Supp. 1987), as are methods for constructing such cell lines (see *e.g.*, Ausubel *et al.*, *supra*). In one example, cDNA encoding an epitope, or epitope-containing antigen is cloned into an expression vector that includes the dihydrofolate reductase (DHFR) gene. Integration of the plasmid and, therefore, integration of the T cell epitope, or epitope-containing antigen-encoding DNA into the host cell chromosome is selected for by inclusion of 0.01-300 μM methotrexate in the cell culture medium (as described, Ausubel *et al.*, *supra*). This dominant selection can be accomplished in most cell types. Recombinant protein expression can be increased by DHFR-mediated amplification of the transfected gene. Methods for selecting cell lines bearing gene amplifications are described in Ausubel *et al.* (*supra*). These methods generally involve extended culture in medium containing gradually increasing levels of methotrexate. The most commonly used DHFR-containing expression vectors are pCVSEII-DHFR and pAdD26SV(A) (described in Ausubel *et al.*, *supra*). The host cells described above or, preferably, a DHFR-deficient CHO cell line (*e.g.*, CHO DHFR$^-$ cells, ATCC Accession No. CRL 9096) are among those most preferred for DHFR selection of a stably-transfected cell line or DHFR-mediated gene amplification.

Another preferred eukaryotic expression system is the baculovirus system. One baculovirus expression system that is commercially available is the cloning vector pBacPAK9, which is available from Clontech (Palo Alto, CA). If desired, this system may be used in conjunction with other protein expression techniques, for

example, the myc tag approach described by Evan *et al.* (Mol. Cell Biol. 5:3610-3616, 1985).

Once the recombinant protein is expressed, it can be isolated from the expressing cells by cell lysis followed by protein purification techniques, such as affinity chromatography. Once isolated, the recombinant protein can, if desired, be purified further by *e.g.*, by high performance liquid chromatography (HPLC; *e.g.*, see Fisher, Laboratory Techniques In Biochemistry And Molecular Biology, Work and Burdon, Eds., Elsevier, 1980).

Polypeptides of the invention, particularly short, independent T cell epitopes, can also be produced by chemical synthesis (*e.g.*, by the methods described in Solid Phase Peptide Synthesis, 2nd ed., 1984, The Pierce Chemical Co., Rockford, IL). These general techniques of polypeptide expression and purification can also be used to produce and isolate immunodominant T cell epitopes for as described herein.

Those skilled in the art of molecular biology will understand that a wide variety of expression systems may be used to produce the recombinant T cell epitope, or epitope-containing antigen. The precise host cell used is not critical to the invention. The epitope, or epitope-containing antigen may be produced in a prokaryotic host (*e.g.*, *E. coli*) or in a eukaryotic host (*e.g.*, *S. cerevisiae*, insect cells such as Sf9 cells, or mammalian cells such as COS-1, NIH 3T3, or HeLa cells). These cells are commercially available from, for example, the American Type Culture Collection, Rockville, MD (see also Ausubel *et al.*, *supra*). The method of transformation and the choice of expression vehicle (*e.g.*, expression vector) will depend on the host system selected. Transformation and transfection methods are described, *e.g.*, in Ausubel *et al.* (*supra*), and expression vehicles may be chosen from those provided, *e.g.*, in Pouwels *et al.*, *supra*.

Alternate Routes of Administration of T Cell Epitopes Derived From or Contained Within an Antigen

In the following example, HIV gp120 is modified in an effort to create an anti-

HIV vaccine, and the modified rgp120 is administered with an adjuvant (ISCOM) intramuscularly. This is but one administration protocol; any adjuvant may be used, and the rgp120, or any other T cell epitope derived from an antigen, or an antigen engineered to have enhanced T cell epitopes according to the methods described herein, may be administered without an adjuvant if so desired. Additionally, more than one epitope may be administered simultaneously. For example, if there are two immunodominant epitopes of an antigen, or more than one antigen, both epitopes may be administered in concert. It will also be understood, that a T cell epitope derived from an antigen, or an antigen engineered to have enhanced T cell epitopes, may be administered within any pharmaceutically-acceptable diluent, carrier, or excipient, in unit dosage form. Methods of administration may differ depending upon whether the T cell epitope is being used as a vaccine, an immunostimulant for an already infected individual, or as a tolerogen for an individual suffering from an autoimmune disease. Conventional pharmaceutical practice may be employed to provide suitable formulations or compositions to administer T cell epitopes derived from an antigen, or antigens engineered to have enhanced T cell epitopes to patients suffering from a disease (*e.g.*, virally-induced hepatitis) that is caused by an insufficient immune response, or to healthy individuals as a vaccine to prevent onset of a disease caused by an insufficient immune response. Administration may begin before an infected patient is symptomatic. Any appropriate route of administration may be employed, for example, administration may be parenteral, intravenous, intra-arterial, subcutaneous, intramuscular, intracranial, intravaginal, intraorbital, ophthalmic, intraventricular, intracapsular, intraspinal, intracisternal, intraperitoneal, intranasal, aerosol, by suppositories, or oral administration. Therapeutic formulations may be in the form of liquid solutions or suspensions; for oral administration, formulations may be in the form of tablets or capsules; and for intranasal formulations, in the form of powders, nasal drops, or aerosols.

Methods well known in the art for making formulations are found, for example, in Remington's Pharmaceutical Sciences, (18th edition), ed. A. Gennaro,

1990, Mack Publishing Company, Easton, PA. Formulations for parenteral administration may, for example, contain excipients, sterile water, or saline, polyalkylene glycols such as polyethylene glycol, oils of vegetable origin, or hydrogenated napthalenes. Biocompatible, biodegradable lactide polymer, lactide/glycolide copolymer, or polyoxyethylene-polyoxypropylene copolymers may be used to control the release of the compounds. Other potentially useful parenteral delivery systems for T cell epitopes derived from an antigen, or antigens engineered to have enhanced T cell epitopes include ethylene-vinyl acetate copolymer particles, osmotic pumps, implantable infusion systems, and liposomes. Formulations for inhalation may contain excipients, for example, lactose, or may be aqueous solutions containing, for example, polyoxyethylene-9-lauryl ether, glycocholate and deoxycholate, or may be oily solutions for administration in the form of nasal drops, or as a gel.

If desired, treatment with a T cell epitope derived from an antigen, or an antigen engineered to have enhanced T cell epitopes may be combined with more traditional therapies for the specific disease, such as surgery, steroid therapy, or chemotherapy for autoimmune disease; antiviral therapy for AIDS or other viral infections; and radiotherapy, chemotherapy, or surgery for cancer.

HIV gp120: A model case for enhanced immunogenic T cell epitopes

In a model case utilizing my T cell epitope detecting methods and application of these methods for the production of improved T cell epitopes (by inserting unstable polypeptide segments into an antigen), I have focused my efforts on the gp120 antigen. Gp120 is produced by HIV, the etiological agent of Acquired Immunodeficiency Syndrome (AIDS). Utilizing the methods of the invention, a vaccine that confers protection against HIV infection may be developed. It will be clearly understood that although the following experiments pertain to HIV gp120, the methods of the present invention may be used to manipulate any number of antigens from any number of sources, be they viral, bacterial, or eukaryotic in origin.

In a normal immune response to a virus, $CD4^+$ T lymphocytes typically activate cytotoxic $CD8^+$ T lymphocytes, which then selectively kill virus-infected cells. Unfortunately, very little is known about what constitutes an effective $CD4^+$ T lymphocyte response against HIV. Very few individuals mount protective immune responses against HIV following infection: most HIV infected patients eventually develop the fatal disease AIDS.

Despite the large number of HIV infected patients world-wide, there is currently no rational methodology for modifying or augmenting antigen presentation of HIV peptides to $CD4^+$ T lymphocytes. The invention described herein will allow the augmentation of the $CD4^+$ T cell response against HIV gp120 through the engineering of enhanced T cell epitopes within gp120. This will allow the development and use of vaccines which will stimulate an anti-HIV $CD4^+$ T cell response, even in individuals already infected.

Insertion of unstable polypeptide segments in HIV gp120

A) Selection of human Hsp10 sequence to be inserted

The mobile loop of human Hsp10 is ideally suited for use as a proteolytically sensitive insert that directs presentation of a flanking epitope in gp120. The mobile loop of the *M. leprae* Hsp10 (*M. leprae* cpn10; see Fig 1A) is associated with a C-terminal immunodominant T-cell epitope, and its immunodominance is likely to derive from enhanced proteolytic sensitivity in the adjacent loop. I have previously demonstrated enhanced sensitivity in mobile loops of two Hsp10 proteins (Landry *et al.*, Nature 364: 255-258, 1993; Landry *et al.*, Proc. Natl. Acad. Sci. USA 93: 11622-11627, 1996). Proteolytic sensitivity may be attributed to the loop's flexible disorder, and this is a conserved feature of all Hsp10s so far examined. Insertion of a mobile loop should not significantly disturb the three-dimensional structure of gp120. Loop disorder in the native Hsp10 indicates a lack of conformational restraint within the loop, and thus it is not likely to distort flanking sequences. Since the ends of the loop lie close together in Hsp10, insertion of the loop into gp120 should mimic its native

disposition, and therefore its native conformational flexibility should be preserved.

The human Hsp10 mobile loop was selected over loops from other Hsp10s because it is likely to be poorly immunogenic. Human and mouse Hsp10s elicit a very weak antibody response in mice, and this may due to tolerance. Thus, a strong antibody response is not expected to be generated in response to the mobile loop insert in gp120. An antibody response may be an issue for concern since antibody binding to antigen could affect antigen processing and presentation (Manca *et al.*, J. Immunol. 9: 2893-2898, 1988). Our studies indicating that flexible segments are antigen processing sites have raised the possibility that antibody binding to the flexible segments could block processing. Ironically, a good antibody response to gp120 might lead to poor cellular responses later in the infection (or boosting schedule). Likewise, T-cell epitope "spreading" in autoimmune disorders has been explained by changes in antigen processing caused by antigen-specific antibodies (Holmdahl *et al.*, Immunol. Rev. 144: 109-135, 1995; Mamula, M. J., Immunol. Rev. 144: 301-315, 1995). Our use of the mammalian Hsp10 mobile loop may allow a level of B-cell tolerance which helps avoid affects on processing caused by antibody binding.

B) Selection of sites in gp120 for loop insertion

The sites for loop insertion in gp120 were chosen with the goal of enhancing the T-cell response to epitopes in gp120 that are likely to be recalled upon HIV infection. The loop insertions also should not severely disturb potentially neutralizing antibody epitopes in gp120.

Sites in gp120 were selected for insertion of flexible loops by the following preferred criteria:

1. The site should lie on the N-terminal side of a sequence that has characteristics of a good T-cell epitope:

   i) the sequence should be conserved among HIV variants;

   ii) the sequence should contain hydrophobic residues that will serve as anchors for binding in the MHC cleft; and

iii) the sequence must not contain cysteines. A disulfide bond could restrict the availability of the target epitopes.

2. The target T-cell epitope should be among epitopes primed in HIV infections.

3. The site should not be located in a known neutralizing antibody epitope nor in a sequence where mutations are known to disturb a neutralizing antibody epitope.

Three sites in gp120 were selected for insertion of the human Hsp10 mobile loop (sites indicated by arrows in Fig. 15). Each site is in a different conserved region of gp120, and each lies within or N-terminal to an immunodominant region of gp120 as defined by cellular proliferative responses. The predicted structural context and the potential damage to neutralizing antibody epitopes was also considered. Each insertion site is in a sequence predicted to form a turn or make a structural transition (Fig. 15). This will allow the loop to be exposed while causing the least disturbance to the gp120 structure. Residue 81 lies in the C1 region immediately after a proline-rich sequence. This site is N-terminal to a marginally solvent-exposed segment followed by a buried immunodominant region. No sites were selected in the regions V1, V2, C2, and V3 since these regions are associated with neutralizing antibody epitopes, highly variable, and not very immunogenic with respect to T-cell epitopes. Residue 334 lies just C-terminal to one of the cysteines that closes the V3 loop, in a segment of the C3 region for which no antibody specificity has been assigned and thus is likely to be buried. This site lies in the middle of an immunodominant region spanning the V3 loop and part of the C3 domain. No sites were selected in the V4 or C4 domains since they are involved in binding to CD4 and therefore likely to present neutralizing antibody determinants. Residue 468 lies in a glycine-rich sequence of the V5 domain, in a marginally solvent accessible segment N-terminal to a sequence rich in T-cell epitopes.

Production of wild-type and engineered recombinant gp120 (rgp120)

The gp120 proteins (wild-type and the three variants) are expressed in insect cells using the baculovirus expression system and purified by lectin affinity

chromatography. Furthermore, since envelope glycoproteins expressed in mammalian cells are better immunogens with respect to eliciting antibodies against conformational epitopes, development of stably transfected cell lines expressing rgp120 proteins may also be employed. Transfection of plasmid DNA encoding rgp120 proteins into a mammalian cell may be by any of the various methods (*e.g.*, DEAE-dextran, $CaPO_4$ precipitation, electroporation) well known in the art of molecular biology (see Ausubel *et al.*, *supra*). Mammalian cells stably incorporating this exogenous DNA may be generated by co-transfecting the rgp120 protein-encoding DNA with DNA capable of expressing a eukaryotic selectable marker (*e.g.*, the neo gene). Stable cell lines may then be generated by limiting dilution of G418-resistant transfected cell. Recombinant gp120 proteins expressed by eukaryotic cells may also be purified by lectin affinity chromatography.

In a preferred example, the rgp120 proteins are expressed by baculovirus in insect cells and purified by lectin affinity chromatography. Recombinant baculovirus expressing the HIV-1/89.6 gp120 is prepared by recombination of baculovirus DNA with a transfer vector containing the HIV-1/89.6 gp120 sequence, according to standard methods (see Ausubel *et al.*, *supra*). The transfer vector will be prepared by replacing the gp120 sequence in the HXBc2 transfer clone (Lu *et al.*, J. AIDS Hum. Retrovir. 12: 99-106, 1996). Expression of rgp120 proteins will be carried out in Hi-Five cells (commercially available from Invitrogen, Carlsbad, CA). The rgp120 will be adsorbed to immobilized Galanthus nivalis agglutinin and eluted with methyl-alpha-D-mannopyranoside, as previously described (Gilljam, G., AIDS Res. And Hum. Retrovir. 9: 431-438, 1993). Using this method, 2 mg of protein per liter of culture is obtained.

A sequence encoding the human cpn10 mobile loop is inserted into the gp120 transfer vector after modification (*e.g.*, by PCR amplification) to introduce unique restriction sites at the positions indicated above (see arrows in Fig. 15). The mobile loop sequence is amplified from the Hsp10 cDNA using primers containing restriction sites compatible with those engineered in the gp120 sequence.

It will be understood that processing site insertions may be incorporated into more native-like immunogens. Likewise, this strategy for the enhancement of $CD4^+$ T cell epitopes may be useful in other vaccine formulations, including DNA and virus-based immunogens.

Preparation of rgp120 with the adjuvant, ISCOM

ISCOMs (immunostimulating complexes) were selected as an adjuvant system since they have been reported to induce strong TH1 as well as TH2 responses (Kersten and Crommelin, Biophys Acta 1241: 117-138, 1995; Barr and Mitchell, Immunol. Cell. Biol. 74: 8-25, 1996; Tarpey *et al.*, Vaccine 14: 230-236, 1996; Cox and Coulter, Vaccine 15: 248-256, 1997). Additionally, I will use purified saponin, QH-C, for formation of ISCOMs to minimize any toxicity associated with crude saponin preparations.

Preparation of protein micelles, and subsequent formation of ISCOMs is performed by the dialysis method as previously described (Lovgren *et al.*, J. Immunol. Methods 98: 137-143, 1987; Blomstedt *et al.*, J. AIDS Res. Hum. Retrovir. 12: 213-220, 1996). Briefly, 2 mg of native rgp120 or each containing flexible regions rgp120 (or keyhole limpet hemocyanin, as a control) is dissolved in PBS containing 2% MEGA-10. These protein micelles are added to the ISCOM formulation mixture consisting of 10 mg of cholesterol and 10 mg of phosphatidyl choline in 20% MEGA-10, followed by addition of 10 mg of the purified Quillaja saponin, QH-C (commercially availalbe from Isotec AB, Inc.). Following sonication, the mixture is left at 25°C for 60 min and then is dialyzed versus PBS overnight. The dialyzed sample is then centrifuged (200,000 x g) on a 20% sucrose cushion to pellet the rgp120 containing ISCOMs. The amount of rgp120 incorporated into each adjuvant preparation is determined by visual inspection of SDS-PAGE resolved proteins. After determination of protein content, the ISCOM-rgp120 preparations are stored at -70°C in aliquots for use as immunogens, as has been previously described (Sjolander *et al.*, Vaccine 14: 344-352, 1996).

Immunization of Rhesus monkeys with rgp120

Once the recombinant HIV gp120 constructs have been made, five groups of of two age-matched, male Rhesus macaques will be immunized as indicated below:

Group 1: 200 µg of keyhole limpet hemocyanin formulated in ISCOMs.

Group 2: 200 µg of native, recombinant gp120 formulated in ISCOMs.

Group 3: 200 µg of gp120 containing a flexible insert at position 80 (rgp120i80) formulated in ISCOMs.

Group 4: 200 µg of gp120 containing a flexible insert at position 335 (rgp120i335) formulated in ISCOMs.

Group 5: 200 µg of gp120 containing a flexible insert at position 469 (rgp120i469) formulated in ISCOMs.

Since single immunizations with HIV gp120 have not resulted in high levels of antibody production, it will be necessary to immunize multiple times. Hence, the animals will receive three immunizations, each 3 months apart from each other.

The first two immunizations will be followed one month later by collection of sera to determine anti-rgp120 antibody levels in each animal. These assays are performed essentially to follow the development of the antibody response against HIV gp120 to assure that the immunizations are effective.

ELISA to detect antibody against rgp120 and peptides derived from gp120

Antibody titers against rgp120 proteins and peptides thereof are determined as previously described (Pascual and Bost, Peptide Res. 2: 207-212, 1989; Pascual and Bost, Immunol. Invest. 19: 421-433, 1990; Takahashi *et al.*, Inf. and Imm. 64: 12990-1298, 1996; Takahashi *et al.*, J. Infect. Dis. 173: 627-635, 1996). Briefly, microtiter plates are coated with rgp120 protein, followed by blocking the plates with BSA. Sera is diluted in PBS containing 1% BSA to determine antibody titers. Serial 1:3 dilutions beginning with a 1:1000 dilution are used, and bound antibody detected using an HRP conjugated anti-human IgG (γ chain, Southern Biotechnology), followed by addition of substrate (TMB).

For peptide mapping studies, 48 peptides representing 20mers with

overlapping 10 amino acid segments, derived from rgp120 (strain 89.6), will be used. Two different dilutions of each sera (1:25 and 1:250) will be used to assure that reactivities can be accurately compared.

Analysis of the anti-gp120 $CD4^+$ T cell response in immunized Rhesus monkeys

Following the third immunization, a detailed analysis of the antigen-specific $CD4^+$ T lymphocyte response is performed.

Phase 1: The $CD4^+$ T lymphocyte response in each immunized primate using overlapping, synthetic peptides derived from HIV gp120 strain 89.6. Since 48 overlapping 20 amino acid peptides are required to completely map gp120, peptide pools are used initially to stimulate isolated $CD4^+$ T lymphocytes. Autologous B lymphoblastoid cell lines (BLCL) are used to present the peptides. In addition, both TH1 (IL-2 and IFN-γ) and TH2 (IL-4 and IL-10) cytokine mRNA expression levels are determined.

Establishment of B lymphoblastoid cell lines (BLCL) for use as antigen presenting cells

A Rhesus Epstein-Barr virus (RhEBV,224) is used to transform autologous B lymphocytes from each of the 10 primates prior to immunization of these animals with rgp120, following the methods previously described (Ohkawa *et al.*, AIDS Res. Hum. Retrovi. 10: 27-38, 1994). Peripheral blood mononuclear cells are isolated from each primate by centrifugation on hypaque-ficoll gradients. 10^7 mononuclear leukocytes are cultured in 2 ml of RPMI-1640 containing 10% heat-inactivated fetal calf serum (FCS) with 20 plaque forming units (pfu) of RhEBV per ml for 2 hours at 37°C. Cells are then diluted 5 fold in media-FCS containing 0.5 μg/ml cyclosporin A, and cultured in 25 cm^2 flasks. (To augment transformation of B lymphocytes, I have found that it is optimal to inhibit T lymphocyte function by the addition of cyclosporin A.) Flasks are monitored for several weeks to observe cell growth. Expansion of cells is followed by cryopreservation of aliquots of cells. These autologous B lymphocytic cell lines (BLCL) derived for each Rhesus monkey will be

used as antigen presenting cells for $CD4^+$ T cell responses as well as for CTL assays described below.

Stimulation of $CD4^+$ T lymphocytes with peptides derived from HIV gp120 (strain 89.6)

For class II-restricted antigen presentation, autologous B lymphoblastoid cell lines (BLCL) for each primate will be treated as described for stimulation of $CD4^+$ T lymphocytes (Monji and Pious, J. Immunol. 158: 3155-3164, 1997). Briefly, BLCL are expanded, harvested, and washed in PBS. Cells are resuspended in 0.005% glutaraldehyde for 30 seconds at 25°C, followed by addition of 0.2 M L-lysine, and two washes in PBS. Per well of a 24 well tissue culture plate, 10 μM of each synthetic peptide derived from HIV gp120 (strain 86.9) will be added to 2 X 10^6 fixed BLCL in complete media (RPMI-1640 supplemented with 4mM L-glutamine and gentamicin) with 10% heat-inactivated normal human AB sera (ABI, Columbia, MD). This is done just prior to the start of isolating $CD4^+$ T lymphocytes (see below). Thus, the antigen presenting cells have about a three hour pre-incubation with the gp120 peptides prior to addition of the autologous $CD4^+$ T lymphocytes.

Isolation of $CD4^+$ T lymphocytes

Peripheral blood mononuclear cells are isolated from 15 ml of Rhesus blood by centrifugation on hypaque-ficoll gradients. $CD4^+$ T lymphocytes are then negatively selected using a two step process. First, the mononuclear leukocytes are incubated for 30 minutes at 4°C on polyclonal anti-human IgM (heavy and light chain, Southern Biotechnology Associates, Birmingham, AL) coated plates. This incubation not only specifically removes B lymphocytes, but also results in the removal of monocytes via non-specific adherence to the plastic plates. Mononuclear cells not adhering the anti-Ig coated plates are harvested, centrifuged and resuspended in media containing magnetic anti-CD8 conjugated microbeads (commercially available from Miltenyi Biotec, Auburn, CA). These magnetic beads are conjugated with the monoclonal anti-CD8 antibody (Leu 2A) which recognizes Rhesus CD8 (commercially available from Becton Dickinson). After 30 minutes, mononuclear

cells not adhering to the anti-CD8 magnetic beads are removed, centrifuged, and counted. 10^6 $CD4^+$ T lymphocytes per well are then added to the 24 well plates which already contain fixed, autologous BLCL and peptide antigen. After three days of stimulation, RNA is isolated from each well to determine lymphokine mRNA profiles using semi-quantitative RT-PCR (see below). For lymphokine secretion, 100 µl of supernatant is taken at days 4 and 6 post stimulation for quantification by ELISA (also below).

Phase 2: The antigen-specific TH1 and TH2 mRNA responses of $CD4^+$ T lymphocytes from immunized Rhesus macaques are more accurately and quantitatively using the peptides defined in Phase 1. Specifically, for those pools of peptides which augment the TH1 or TH2 responses, a determination of which peptides in each pool were recognized is determined. Once again the readout will be RT-PCR for TH1 and TH2 lymphokine mRNA expression.

Phase 3: Increased mRNA expression observed in Phase 2 translates into secretion of these lymphokines. For those HIV gp120 peptides which stimulate the most significant differences in lymphokine mRNA expression between groups of immunized macaques, ELISAs to quantify secretion of that particular lymphokine are performed.

Longitudinal studies will also be performed by a fourth immunization of each Rhesus macaque with the appropriate immunogen 11 months after the third immunization (*i.e.*, 17 months after the first immunization). The magnitude, specificity and lymphokine phenotype of antigen-specific $CD4^+$ T lymphocytes are defined a month after the fourth immunization, and again 12 months later (*i.e.*, 29 months after the first immunization). This will allow an evaluation of the relative responsiveness of memory $CD4^+$ T lymphocytes generated by immunization with each rgp120 construct. In addition, mapping studies allows the determination of whether or not the specificity of the $CD4^+$ T lymphocyte response changes with time.

From these results, significant alterations in the response of antigen-specific $CD4^+$ T cells from Rhesus macaques immunized with rg120 containing flexible

segments when compared to macaques immunized with native rgp120 may be found. Furthermore, significant differences between groups where the flexible segments are in different locations may also found. Specifically, epitopes flanking the inserted flexible regions become the immunodominant $CD4^+$ T cell antigens in the respective groups, although the $CD4^+$ T cell response against the remainder of the gp120 protein may remain intact.

$CD4^+$ T cell responses in HIV infection: a focus on TH1/TH2-derived cytokines

There are at least two different functional subsets of $CD4^+$ helper T lymphocytes (TH) which can be identified based on their ability to secrete particular lymphokines. Antigen-specific TH1 lymphocytes secrete substantial amounts of interleukin-2 (IL-2) and interferon-γ (IFN-γ), and augment cell-mediated immune responses (*e.g.*, cytotoxic T lymphocyte activity and macrophage activation). Conversely, antigen-specific TH2 lymphocytes secrete IL-4, IL-5, IL-6, IL-10, and IL-13, and augment antibody production by B lymphocytes. Cytokine production which favors development of TH1 responses also serves to limit development of TH2 responses, and vice versa. Hence, stimulation of one T helper cell subset and not the other results in the production of a particular set of cytokines which may define the resulting immune response.

There is evidence that infection with HIV results in an alteration of the $CD4^+$ T lymphocyte response toward a TH2 phenotype. A suggestion that HIV-infected individuals underwent a progressive TH1 to TH2 shift helped explain the loss of T lymphocyte reactivity to recall antigens in these patients (Clerici and Shearer, Immunol. Today 14, 107-111, 1993; Shearer and Clerici, Immun. Today 15: 575-581, 1994). In addition, more recent reports also suggest a shift towards development of TH2 responses in these patients (Diaz-Mitoma *et al.*, Clin. Exp. Immunol. 102: 31-39, 1995; Hyjek *et al.*, J. Immunol. 155: 4060-4071, 1995; Jason *et al.*, J. Acquir. Defic. Syndr. Hum. Retrovirol. 10: 471-476, 1995; Meyaard *et al.*, J. Immunol 157: 2712-

2718, 1996). If T cell epitopes can be found to be associated with a TH1 or a TH2 immune response, then the present invention, which allows one to rationally enhance TH1 and/or TH2 responses against HIV gp120, provides a significant advantage for vaccine development in HIV infected patients, as well as in non-infected individuals.

Following administration to Rhesus monkeys of HIV gp120 and gp120 engineered to bear enhanced T cell epitopes, an analysis is made of antigen-specific $CD4^+$ T cell responses based upon TH1/Th2 lymphokine mRNA expression and lymphokine secretion. Analyses of the antigen-specific CD4+ T lymphocyte responses may be performed, in this example, in three phases:

Phase 1: Twelve different pools of 4 consecutive, overlapping peptides derived from gp120 are used to stimulate $CD4^+$ T cells from each monkey previously administered gp120 and variants thereof. mRNA lymphokine profiles will be performed on these cells. RNA extracted from antigen-stimulated $CD4^+$ T lymphocytes is reverse transcribed using random hexamers as previously described (Bost, K. L., J. Neuroimmunol. 62L 59-67, 1995; Bost and Clements, Inf. and Imm. 63: 1076-1083, 1995; Bost and Mason, J. Immunol. 155: 285-296, 1995; Bost *et al.*, J. Immunol. 154: 718-729, 1995; Bost *et al.*, Immunol. 87: 633-641, 1996). Semi-quantitative RT-PCR is performed on each RNA sample for expression of TH1 (IL-2 and IFN-γ) and TH2 (IL-4, and IL-10) mRNAs using PCR primers specific for these respective cytokines. RT-PCR for the housekeeping gene, G3PDH is also performed on each sample to control for RNA loading.

Phase 2: For those pools of peptides which stimulate the most significant differences in cytokine mRNA expression between native gp120 immunized primates and those immunized with rgp120 containing flexible segments, a second set of antigen-stimulations is performed. Here the individual peptides making up the peptide pool are used to obtain a more accurate map of peptide-stimulated lymphokine mRNA expression, as described herein. Focus is on the peptide pools which result in the most significant differences in stimulating $CD4^+$ T lymphocyte mRNA expression in animals immunized with native rgp120 versus the other three

groups immunized with rgp120 containing flexible regions. For example, if peptide pool #2 did not stimulate significant cytokine mRNA expression in $CD4^+$ lymphocytes of primates immunized with native rpg120, but primates immunized with rgp120i80 did respond to peptide pool #2, such a difference would be a primary focus of my investigations.

Phase 3: For those peptides which stimulate significant increases in TH1 or TH2 lymphokine mRNA expression, secretion of these lymphokines is then quantified using ELISAs as follows:

IL-2 ELISA: A monoclonal anti-human IL-2 antibody (MQ1-17H12; commercially available from Pharmingen, San Diego, CA) is used as a capture antibody and a biotinylated, polyclonal anti-human IL-2 antibody (Pharmingen) will be used as a detection antibody, and the ELISA assay conducted according to standard methodologies (Ausubel *et al.*, *supra*). Both of the anti-human IL-2 reagents cross react with Rhesus IL-2 (Pharmingen Technical data sheet).

IL-4 ELISA: A commercially available anti-human IL-4 capture ELISA (Amersham, Arlington Heights, IL) which crossreacts with Rhesus IL-4 (Villinger *et al.*, J. Immunol. 155: 3946-3954, 1995) will be used.

IL-10 ELISA: A commercially available anti-human IL-10 capture ELISA (R&D Systems, Minneapolis, MN) which crossreacts with Rhesus IL-10, and does not recognize Epstein Barr-produced viral IL-10 (R&D Systems technical services report) will be used.

IFN-γ ELISA: A commercially available anti-human IFN-γ ELISA (Genzyme, Cambridge, MA) which crossreacts with Rhesus IFN-γ (Genzyme technical bulletin) will be used.

The anti-IL-2 antibody pair was tested on culture supernatants from mitogen stimulated, Rhesus peripheral blood mononuclear cells. There was significant crossreactivity, confirming the findings of the Pharmingen technical support staff. In addition, since previous investigations have demonstrated that using the Amersham capture ELISA to quantify human IL-4 that Rhesus IL-4 crossreacted essentially

completely with the anti-human reagents (Villinger *et al.*, *supra*), it is therefore legitimate to use human IL-4 for standard curves to quantify Rhesus IL-4 secretion. However, this may not be the case for Rhesus IL-2, IL-10 and IFN-γ even though the antibodies against human lymphokines are known crossreact with Rhesus lymphokines. To assure that our standard curves are appropriate for extrapolating values, Rhesus IL-2, IL-10 and IFN-γ will be cloned, expressed, and purified using the pFLAG expression system as previously described (Bost *et al.*, Immunol. 87: 633-641, 1996).

In one example, peptide pools #1, 3, 5, etc. do not stimulate significant cytokine mRNA expression in any of the immunized animals. Hence, there is no need to further map these regions of gp120 or to use these peptides to stimulate lymphokine secretion. Second, phase 2 and phase 3 investigations will focus on differences in antigen-specific $CD4^+$ lymphokine expression between those animals immunized with native rgp120 versus those immunized with the rgp120 proteins containing flexible regions (*i.e.*, rgp120i80, rgp120i335, and rgp120i469). For example, if immunization with all 4 rgp120 constructs result in identical lymphokine expression stimulated by particular peptide pools (*e.g.*, pools 10, 25, 46, etc.), these similarities are not focused upon in phase 2 and phase 3 investigations. Again, the results obtained in phase 1 studies will quickly limit what is done in phase 2 and phase 3. Finally, when increased mRNA expression correlates with increased cytokine secretion, such complementary information using two completely different techniques gives added support for the conclusions to be gleaned from the data. Alternatively, when increased mRNA expression does not correlate with secretion and potential technical problems are eliminated, such results provide very meaningful information concerning post-transcriptional regulation of cytokine secretion. Therefore, taken together the mRNA and secretion analyses are complementary.

Vaccines targeting HIV gp120 : the ability of a subunit HIV gp120 vaccines to stimulate humoral and CTL responses

A subunit vaccine like HIV gp120 may be used to stimulate humoral or CTL responses. The type of immune response against this subunit vaccine candidate which ultimately dominates can depend upon a variety of factors (reviewed in Haynes, B. F., Lancet 348: 933-937, 1996; McElrath *et al.*, AIDS Res. Hum. Retrovir. 13: 211-216, 1997). However, the present invention allows an exploration into the possibility that by enhancing presentation of HIV gp120 peptides to $CD4^+$ T cells epitopes, it will be possible to augment the immune response since $CD4^+$ T cell help is necessary for optimal humoral immunity and CTL responses.

Antibody-mediated neutralization of SHIV virus infectivity in vitro

For propagation of virus, CEMx174 cells (publically available from the NIH AIDS Research and Reference Reagent Program) are transfected with the ligation of p5'SHIV with p3'89.6u+SHIV, according to methods previously described (Li *et al.*, J. Virol 69: 7061-7067,1992). Virus stocks are prepared by infecting Rhesus monkey peripheral blood mononuclear cells. Viral titers are determined from cell-free supernatants using serial dilutions of virus containing supernatants placed on CEMx174 cells. The 50% tissue culture infectious dose (TCID50) is calculated from microscopic inspection of syncytia formation after 2 weeks.

Antibody neutralization is performed essentially as described (Lu *et al.*, J. AIDS and Hum. Retrovir. 12: 99-106, 1996) with the following modifications: serial dilutions (1:3, beginning at a sera dilution of 1:30) of heat inactivated sera from each Rhesus monkey is mixed with 200 TCID50 of SHIV at 37°C for 60 min. 10^5 CEMx174 cells are then added to each well, and four days viral production is measured by quantifying SIV p27 antigen release into the culture supernatant (Coulter Co,. Miami, FL). Neutralizing titers are defined as the last dilution to provide 50% protection when compared SIV p27 levels in control antibody treated wells challenged with virus.

CTL assays

Cytotoxic T lymphocyte (CTL) responses are performed as previously described (Voss *et al.*, Virol. 208: 770-775, 1995). Peripheral blood mononuclear cells are isolated and cultured for three days in RPMI 1640 containing 10% FCS and gentamicin with 2 μg/ml of concanavalin A. After three days, the cells are expanded into media containing 100 U/ml recombinant human IL-2 for an additional two days.

For targets, autologous B lymphoblastoid cell lines (BLCL) established for each Rhesus monkey as described above are infected with the recombinant vaccinia virus, vBD3 (Dr. Robert Doms, University of Pennsylvania), which expresses a full length HIV envelope protein (strain 89.6) under control of the vaccinia virus early/late promoter. Two plaque forming units per ml of vBD3 are added to target cells for 16 hours, after which they are loaded with ^{51}Cr, according to standard protocols (see Ausubel *et al.*, *supra*). Peripheral blood mononuclear leukocytes are isolated from each macaque following the third or fourth immunization, and used as effector cells. Four different effector to target cell ratios will be used (50:1, 25:1, 12:1 and 6:1). Percent specific lysis is determined using background lysis as that observed in macaques immunized with keyhole limpet hemocyanin.

For those macaques which demonstrate significant CTL activity against autologous BLCL infected with vaccinia expressing gp120, overlapping synthetic peptides are employed to map the response. For these studies, autologous BLCLs are pulsed with pools of four consecutive peptides each (*e.g.*, pool #1 = peptides 1-4) for 2 hours prior to addition of mononuclear leukocytes at a ratio of 25:1. Percent specific lysis is determined for each peptide pool using background lysis as that observed in control macaques immunized with keyhole limpet hemocyanin.

For those pools of peptides which are recognized by CTLs, addition experiments are performed to determine which peptides within the pool are being recognized. The assay will be performed as outlined above.

Augmented humoral antibody responses against selected epitopes within HIVgp120 are found when comparing between groups of immunized macaques. Specifically, insertion of flexible regions into rgp120 immunogens is found to

augment and focus the $CD4^+$ T cell response to particular epitopes, and this in turn focuses T cell help. The result of such manipulation is an augmentation and focusing of the humoral immune response.

An augmented CTL response against selected epitopes of HIV gp120 is also found when comparing between groups of immunized macaques. Specifically, insertion of flexible regions into rgp120 immunogens augment and focus the $CD4^+$ T cell response to particular epitopes, and this in turn focuses T cell help. The result of such manipulation is an augmentation and focusing of the CTL response as well.

Testing an rgp120 anti-HIV vaccine in the SHIV primate model for HIV

To conduct these studies, the SHIV primate model for HIV is utilized. Simian/human immunodeficiency viruses (SHIVs) were developed specifically to overcome the shortcomings of SIV for HIV drug and vaccine development (Li *et al.*, J. AIDS 5: 639-646, 1992). SHIVs combine the gag and pol genes of SIV with the env gene of HIV. A new SHIV encoding the envelope glycoprotein of the macrophage tropic HIV strain designated HIV-1/89.6 that is particularly cytopathic in human cells was recently developed (Reimann *et al.*, J. Virol 70: 3198-3206, 1996; Reimann *et al.*, J. Virol. 70: 6922-6928, 1996). The close similarity of SHIV-89.6 infection to rapidly progressing HIV infection coupled with the potential for testing anti-HIV envelope glyoprotein immune responses makes the SHIV-89.6 an exceptional model system for testing novel vaccines.

Rhesus monkeys pre-immunized with rgp120 are challenged with SHIV infection to assess resistance conferred by the rgp120 immunization. In addition, non-immunized animals may also be infected with SHIV, followed by immunization with rgp120 to assess the effects of the rgp120 vaccine in ameliorating the symptoms of SHIV infection.

Mapping Immunodominant Helper T-cell Epitopes in Bacteriophage Gp31

The ability to predict helper T-cell epitopes on the basis of local antigen stability was next examined with a model antigen, bacteriophage T4 Gp31 (SEQ ID NO: 2; GenBank Accession No. M34502). This protein is homologous to the *M. leprae* GroES in terms of its three-dimensional structure and local stability, but the proteins share less than 20% sequence identity. The pattern of helper T-cell epitope immunodominance in Gp31 was determined by mapping lymphoproliferative responses with overlapping synthetic 15-mer peptides spanning the Gp31 sequence.

Two groups of C57/B6 mice (commercially available from the Jackson Laboratory, Bar Harbor, ME) were immunized as follows:

Group 1. 100 μg Gp31 (full length protein) plus 100 μl Freund's complete adjuvant in a total volume of 200 μl (100 μl intrapertoneal and 100 μl subcutaneous).

Group 2. 100 μg keyhole limpet hemocyanin (KLH) plus 100 μl Freund's complete adjuvant in a total volume of 200 μl (100 μl intrapertoneal and 100 μl subcutaneous).

At 10 and 21 days after immunization, all mice given a boost of 50 μg Gp31 (group 1) or KLH (group 2) with 50 μl Freund's incomplete adjuvant in a total volume of 100 μl intraperitoneally.

At 22 days after the second boost, one mouse of each group was sacrificed, and total splenic cells were obtained. Cells ($2x10^5$) were incubated for several days with 100 μg of each 15-mer peptide and then [^{3}H]thymidine was added to the cell culture media, according to standard methods (see, *e.g.*, Ausubel *et al.*, *supra*; Coligan, J.E., Current Protocols in Immunology, John Wiley & Sons, New York, NY, 1991). Cells were incubated for another day and then harvested. [^{3}H]thymidine incorporation was measured by scintillation counting. Background incorporation was determined for culture incubated with no added peptide. The results shown in Fig. 16A are reported as a Stimulation Index (SI), which is defined as the ratio of background-subtracted radioactivity for cells primed with Gp31 to cells primed with

the irrelevant antigen, KLH.

SI values greater than 1 were observed for seven peptides (represented as black bars in Fig. 16C). All seven were centered in regions of Gp31 characterized by high hydrophobicity (Fig. 16B) and low B-factors (Fig.16D). Based on these results, helper T- epitopes were concluded to be preferentially localized to hydrophobic and well-ordered regions of Gp31 which are likely to be protected against proteolytic digestion during antigen processing.

Detection of Immunity to a Predicted T cell Epitope from an Endogenous Antigen

Although the invention described herein is clearly useful in the generation of improved vaccines to stimulate the immune response against a pathogen or other invasive cell type (*e.g.*, a cancer cell), it is equally useful in the design of tolerogenic peptides to inhibit the immune response. Such inhibitory peptides are useful in the treatment of autoimmune diseases.

A predicted epitope is the basis for design of a tolerogenic epitope that reduces immune responses. In one example, if the epitope is derived from a self antigen, such as insulin, diabetes, or a predisposition to develop diabetes, may be detected in a patient by using the methods of the present invention to isolate an immunodominant epitope of insulin. The patient's T lymphocytes may then be isolated and utilized in a T cell proliferation assay (*e.g.*, ^{3}H-thymidine incorporation assay) following stimulation with the epitope in the context of fixed autologous antigen presenting cells according to the methods described herein, and, additionally, following methods described in Ausubel *et al.*, *supra*. Non-radioactive cell proliferation assays are also known in the art (*e.g.*, the CellTiter 96® AQ_{ueous} Non-Radioactive Cell Proliferation Assay commercially available (Promega Co, Madison, WI). Should the patient prove to have T cells which do respond to insulin, tolerogenic epitopes may be designed which may then be administered to the patient to tolerize insulin-specific T cells, thereby preventing onset of diabetes and/or alleviating diabetic symptoms.

Design of Tolerogens for Treatment of Autoimmune or Allergy Disorders

In establishing tolerance to an antigen, a number of factors plays a key role. For example, the dose of the administered antigen is an important variable. Very high concentrations of soluble proteins have been shown to be required to suppress the antibody response (Dixon and Maurer, J. Exp. Med. 101: 245-257, 1955). It has also been shown that extremely low doses of antigen can also induce tolerance (Mitchison, N. A., Proc. Roy. Soc. Lond. (Biol.) 161: 275-292, 1964). In addition, the route of antigen introduction is a critical factor in tolerance induction. Intravenous or oral administration of antigen generally favors induction of tolerance (Cremer *et al.*, J. Immunol. 131: 2995-3000, 1983; Chase, M., Proc. Soc. Exp. Biol. 61: 257-259, 1946).

In a new approach to induce tolerance by utilizing the methods described herein to identify immunodominant T cell epitopes, such epitopes are defined for the targeted auto-antigen of any number of autoimmune diseases including, without limitation, insulin (diabetes), acetylcholine receptor (Myasthenia gravis), synovial proteins (Rheumatoid arthritis), erythrocyte membrane proteins (autoimmune hemolytic anemia), type IV collagen (Goodpasture's syndrome), and thryoid-stimulating hormone receptor (Graves' disease). Peptide mapping of these auto-antigens may be undertaken to identify the immunodominant T cell epitope recognized by T lymphocytes of patients with autoimmune disease using the methods described herein. Once an immunodominant T cell epitope to an auto-antigen is identified, it may be used to tolerize and/or remove reactive cells from a patient with the autoimmune disease, or with a predisposition to develop the disease. The epitope may also be used to screen individuals who have auto-reactive T cells, but who are, as yet, a-symptomatic.

Depletion of Auto-reactive T lymphocytes

Antigen presenting cells may be generated according to the methods described herein from a patient being screened and/or treated for an autoimmune disease. It will

be understood that such antigen-presenting cells may be expanded *in vitro* to attain large numbers of cells autologous to an individual patient. Following incubation of the antigen-presenting cell with the immunodominant T cell epitope of, for example, the thryoid-stimulating hormone receptor, the blood of a patient with Graves' disease may be collected and passaged over a surface covered antigen-presenting cells presenting the immunodominant T cell epitope of the thyroid-stimulating hormone receptor. T lymphocytes which recognize and bind to the T cell epitope presented on autologous antigen-presenting cells will adhere to the immobilized cell, thus exiting the blood. The thyroid-stimulating hormone receptor-specific T lymphocyte-depleted blood is then returned to the patient. By repeated treatments, T lymphocytes which recognize self-peptide in the context of self-MHC will be eliminated, hence preventing and/or alleviating Graves' disease symptoms.

It will be understood that the blood of an autoimmune patients may be passaged over autologous antigen-presenting cells presenting the immunodominant T cell epitope of the targeted auto-antigen in bulk, in a method akin to dialysis treatment of diabetic patients. By this modification, auto-antigen specific T lymphocytes may be depleted in one, albeit lengthy, treatment.

Using this method, T cells mediating allergic reactions to, for example, pollen, may be similarly removed by incubating autologous antigen presenting cells with the immunodominant T cell epitope of a pollen antigen detected by the methods disclosed herein. By removing these pollen-reactive T cells from the allergic patient, allergy symptoms may be alleviated.

Induction of Tolerance by Anergy Due to Lack of Appropriate Costimulation

A state of immunological unresponsiveness, also known as anergy, can be induced if a T cell is presented a peptide in context with autologous MHC, but without the additional presence of a number of costimulatory molecules on the antigen presenting cell which include, without limitation, B7 (CD80), ICAM-1 (CD54), LFA-3 (CD58), and CD72 (Mueller *et al.*, Ann. Rev. Immunol. 7: 445-480,

1989). Hence, once an immunodominant epitope is identified in, for example, collagen type IV, antigen presenting cells may be isolated from patients with Goodpasture's syndrome (also known as glomerulonephritis). Following expansion *in vitro*, the antigen presenting cells may be incubated with the collagen type IV immunodominant T cell epitope in addition to incubation with antibodies which bind to one or more of the costimulatory molecules on the antigen presenting cell. Such antibodies are widely commercially available. For example, anti-human B7, ICAM-1, LFA-3, and CD72 antibodies are all available from Coulter Co. (Miami, FL). However, more than one, for example, anti-B7 antibody may be used to ensure complete blocking of the B7 antigen on the antigen presenting cell. T cells from a Goodpasture's syndrome patient are then incubated with these costimulation molecule-blocked autologous antigen-presenting cells presenting the immunodominant type IV collagen T cell epitope. Such treatment will tolerize any type IV collagen reactive T cells in the sample. Repeated treatments will allow the complete anergy of all type IV collagen reactive T lymphocytes, thus alleviating and/or preventing Goodpasture's syndrome.

As with depletion, tolerization treatments may be modified into one extended treatment, thereby passaging the entire blood supply of the patient over autologous costimulation molecule-blocked antigen cells presenting type IV collagen. In addition, T cells of patients suffering from allergies (*e.g.*, to pollen) may be similarly anergized with autologous costimulation molecule-blocked antigen presenting cells presenting pollen antigens.

Other Embodiments

All publications mentioned in this specification are herein incorporated by reference to the same extent as if each independent publication was specifically and individually indicated to be incorporated by reference.

While the invention has been described in connection with specific embodiments thereof, it will be understood that it is capable of further modifications

and this application is intended to cover any variations, uses, or adaptations of the invention following, in general, the principles of the invention and including such departures from the present disclosure that come within known or customary practice within the art to which the invention pertains and may be applied to the essential features hereinbefore set forth, and follows in the scope of the appended claims.

What is claimed is:

Claims

1. A method for stimulating an immune response specific toward a naturally-occurring protein in an animal having an immune system including T cells, said method comprising administering to said animal an altered protein or polypeptide fragment thereof derived from said naturally-occurring protein, wherein an unstable polypeptide segment has been inserted by artifice into said altered protein.

2. The method of claim 1, wherein said naturally-occurring protein is from a pathogen.

3. The method of claim 2, wherein said altered protein or polypeptide fragment thereof is administered to said animal to prevent infection of said animal with said pathogen.

4. The method of claim 1, wherein said naturally-occurring protein is from a neoplastic cell,

5. The method of claim 4, wherein said altered protein or polypeptide fragment thereof is administered to said animal to inhibit growth of said neoplastic cell in said animal.

6. The method of claim 1, wherein said altered protein or polypeptide fragment thereof is administered with a pharmaceutically acceptable carrier, an adjuvant or both.

7. The method of claim 1, wherein said animal is a mammal.

8. The method of claim 7, wherein said mammal is a human.

9. A method for increasing the immunogenicity of a naturally-occurring protein, said method comprising inserting by artifice into said naturally-occurring protein an unstable polypeptide segment to produce an altered protein.

10. The method of claim 9, wherein said altered protein or polypeptide fragment thereof is in a vaccine.

11. The method of claim 1 or 9, wherein said unstable polypeptide segment comprises at least twelve amino acid residues.

12. The method of claim 11, wherein not more than 30% of said amino acid residues are selected from the group of amino acid residues consisting of isoleucine, leucine, valine, tyrosine, phenylalanine, tryptophan, threonine, and methionine.

13. The method of claim 1 or 9, wherein said unstable polypeptide segment comprises a polypeptide sequence that is specifically recognized by a protease.

14. The method of claim 1 or 9, wherein said unstable polypeptide segment has an average hydrophobicity value that is lower than the average hydrophobicity value of said altered protein; has a sequence conservation that is lower than a sequence conservation of said altered protein; has an amide protection factor that is lower than 10^4 wherein said altered protein is in a native conformational state; has an average amide protection factor that is lower than the average amide protection factor for said altered protein in a denatured conformational state; has an NMR order parameter (S^2) of less than 0.8; or has an average B-factor value that is higher than the average B-factor value of said altered protein.

15. The method of claim 1 or 9, wherein said altered protein comprises a T cell epitope.

16. The method of claim 15, wherein said unstable polypeptide segment is inserted N-terminally adjacent to said T cell epitope.

17. The method of claim 15, wherein the C - terminal portion of said unstable polypeptide segment overlaps the N - terminal portion of said T cell epitope.

18. The method of claim 15, wherein said T cell epitope has an average hydrophobicity value that is higher than the average hydrophobicity value of said altered protein; has a sequence conservation that is higher than a sequence conservation of said altered protein; has an amide protection factor that is greater than 10^4 wherein said altered protein is in a native conformational state; has an average amide protection factor that is higher than the average amide protection factor for said altered protein in a denatured conformational state; has an NMR order parameter (S^2) of greater than 0.7; or has an average B-factor value that is lower than the average B-factor value of said altered protein.

19. The method of claim 15, wherein at least 30% of the amino acid residues of said T cell epitope are selected from the group of amino acid residues consisting of isoleucine, leucine, valine, tyrosine, phenylalanine, tryptophan, threonine, and methionine.

20. A method for detecting in a substantially pure protein a polypeptide segment that is likely to be a T-cell epitope, said method comprising the steps of:

(a) identifying an unstable polypeptide segment in said protein; and

(b) identifying a second polypeptide segment adjacent to said unstable polypeptide segment in said protein, said second polypeptide segment likely to be a T cell epitope.

21. The method of claim 20, wherein said unstable polypeptide segment

comprises at least twelve amino acid residues.

22. The method of claim 21, wherein not more than 30% of said amino acid residues are selected from the group of amino acid residues consisting of isoleucine, leucine, valine, tyrosine, phenylalanine, tryptophan, threonine, and methionine.

23. The method of claim 20, wherein said unstable polypeptide segment comprises a polypeptide sequence that is specifically recognized by a protease.

24. The method of claim 20, wherein said unstable polypeptide segment has an average hydrophobicity value that is lower than the average hydrophobicity value of said protein; has a sequence conservation that is lower than a sequence conservation of said protein; has an amide protection factor that is lower than 10^4 wherein said protein is in a native conformational state; has an average amide protection factor that is lower than the average amide protection factor for said protein in a denatured conformational state; has an NMR order parameter (S^2) of less than 0.8; or has an average B-factor value that is higher than the average B-factor value of said protein.

25. The method of claim 20, wherein said unstable polypeptide segment is N-terminally adjacent to said second polypeptide segment.

26. The method of claim 20, wherein the C - terminal portion of said unstable polypeptide segment overlaps the N - terminal portion of said second polypeptide segment.

27. The method of claim 20, wherein said second polypeptide segment has an average hydrophobicity value that is higher than the average hydrophobicity value of said protein; has a sequence conservation that is higher than a sequence conservation

of said protein; has an amide protection factor that is greater than 10^4 wherein said protein is in a native conformational state; has an average amide protection factor that is higher than the average amide protection factor for said protein in a denatured conformational state; has an NMR order parameter (S^2) of greater than 0.7; or has an average B-factor value that is lower than the average B-factor value of said protein.

28. The method of claim 20, wherein at least 30% of the amino acid residues of said second polypeptide segment are selected from the group of amino acid residues consisting of isoleucine, leucine, valine, tyrosine, phenylalanine, tryptophan, threonine, and methionine.

29. A method for identifying the most immunogenic protein in a group of proteins, said method comprising identifying the protein comprising the most unstable polypeptide segment in said group of proteins, wherein said identified protein is the most immunogenic protein in said group of proteins.

30. The method of claim 29, wherein said most immunogenic protein is substantially purified and said group of proteins is from a neoplastic cell, a pathogen, a foodstuff, an allergen, or a tissue targeted in an autoimmune disease.

31. The method of claim 29, wherein said unstable polypeptide segment comprises at least twelve amino acid residues.

32. The method of claim 31, wherein not more than 30% of said amino acid residues are selected from the group of amino acid residues consisting of isoleucine, leucine, valine, tyrosine, phenylalanine, tryptophan, threonine, and methionine.

33. The method of claim 29, wherein said most unstable polypeptide segment has the lowest average hydrophobicity value of any unstable polypeptides segments

of said group of proteins; has the lowest sequence conservation of any unstable polypeptide segment of said group of proteins; has the lowest average amide protection factor of any unstable polypeptide segment of said group of proteins wherein said proteins in said group are in a native conformational state; has the lowest average amide protection factor of any unstable polypeptide segment of said proteins wherein said proteins in said group are in a denatured conformational state; has the lowest NMR order parameter (S^2) of any unstable polypeptide segment of said group of proteins; or has the average highest B-factor value of any unstable polypeptide segment of said group of proteins.

34. The method of claim 29, wherein said protein comprises a T cell epitope.

35. The method of claim 34, wherein said most unstable polypeptide segment is N-terminally adjacent to said T cell epitope.

36. The method of claim 34, wherein the C - terminal portion of said most unstable polypeptide segment overlaps the N - terminal portion of said T cell epitope.

37. The method of claim 34, wherein said T cell epitope has an average hydrophobicity value that is higher than the average hydrophobicity value of said protein; has a sequence conservation that is higher than a sequence conservation of said protein; has an amide protection factor that is greater than 10^4 wherein said protein is in a native conformational state; has an average amide protection factor that is higher than the average amide protection factor for said protein in a denatured conformational state; has an NMR order parameter (S^2) of greater than 0.7; or has an average B-factor value that is lower than the average B-factor value of said protein.

38. The method of claim 34, wherein at least 30% of the amino acid residues of said T cell epitope are selected from the group of amino acid residues consisting of

isoleucine, leucine, valine, tyrosine, phenylalanine, tryptophan, threonine, and methionine.

39. A method for treating an animal that has or is at risk for developing an allergic response, said method comprising administering to said animal a protein or polypeptide fragment thereof from an allergen, wherein said protein or polypeptide fragment thereof is identified as comprising the most unstable polypeptide segment in a group of proteins of said allergen.

40. A method for treating an animal that has or is at risk for developing an autoimmune disease, said method comprising administering to said animal a protein or polypeptide fragment thereof from a tissue targeted in said immune disease wherein said protein or polypeptide fragment thereof is identified as comprising the most unstable polypeptide segment in a group of proteins of said tissue targeted in said autoimmune disease.

41. The method of claim 39 or 40, wherein said protein or polypeptide fragment thereof is in a tolerogen.

42. The method of claim 39 or 40, wherein said protein or polypeptide fragment thereof is administered orally or wherein said protein or polypeptide fragment thereof is administered with a pharmaceutically acceptable carrier.

43. The method of claim 39 or 40, wherein said animal is a mammal.

44. The method of claim 43, wherein said mammal is a human.

45. A substantially pure antigen comprising an unstable polypeptide segment inserted by artifice.

46. The antigen of claim 45, wherein said unstable polypeptide segment comprises at least twelve amino acid residues.

47. The antigen of claim 46, wherein not more than 30% of said amino acid residues are selected from the group of amino acid residues consisting of isoleucine, leucine, valine, tyrosine, phenylalanine, tryptophan, threonine, and methionine.

48. The antigen of claim 45, wherein said unstable polypeptide segment comprises a polypeptide sequence that is specifically recognized by a protease.

49. The antigen of claim 45, wherein said unstable polypeptide segment has an average hydrophobicity value that is lower than the average hydrophobicity value of said substantially pure antigen; has a sequence conservation that is lower than a sequence conservation of said substantially pure antigen; has an amide protection factor that is lower than 10^4 wherein said substantially pure antigen is in a native conformational state; has an average amide protection factor that is lower than the average amide protection factor for said substantially pure antigen in a denatured conformational state; has an NMR order parameter (S^2) of less than 0.8; or has an average B-factor value that is higher than the average B-factor value of said substantially pure antigen.

50. The antigen of claim 45, wherein said substantially pure antigen comprises a T cell epitope.

51. The antigen of claim 50, wherein said unstable polypeptide segment is inserted N-terminally adjacent to said T cell epitope.

52. The antigen of claim 50, wherein the C - terminal portion of said unstable polypeptide segment overlaps the N - terminal portion of said T cell epitope.

53. The antigen of claim 50, wherein said T cell epitope has an average hydrophobicity value that is higher than the average hydrophobicity value of said antigen; has a sequence conservation that is higher than a sequence conservation of said antigen; has an amide protection factor that is greater than 10^4 wherein said antigen is in a native conformational state; has an average amide protection factor that is higher than the average amide protection factor for said antigen in a denatured conformational state; has an NMR order parameter (S^2) of greater than 0.7; or has an average B-factor value that is lower than the average B-factor value of said antigen.

54. The antigen of claim 50, wherein at least 30% of the amino acid residues of said T cell epitope are selected from the group of amino acid residues consisting of isoleucine, leucine, valine, tyrosine, phenylalanine, tryptophan, threonine, and methionine.

55. The antigen of claim 45, wherein said antigen is associated with a pharmaceutically acceptable carrier, an adjuvant, or both.

56. A vaccine comprising the antigen of claim 45.

57. A tolerogen comprising the antigen of claim 45.

SEQUENCE LISTING

<110> The Administrators of the Tulane Educational Fund
Landry, Samuel J.

<120> PREDICTION, DETECTION AND DESIGN OF T-CELL EPITOPE

<130> 07005/003WO2

<140> NOT KNOWN
<141> 1998-07-28

<150> 60/054,156
<151> 1997-07-29

<160> 2

<170> FastSEQ for Windows Version 3.0

<210> 1
<211> 513
<212> PRT
<213> Bacteriophage T4

<400> 1

```
Met Arg Val Lys Gly Ile Arg Lys Asn Tyr Gln His Leu Trp Lys Trp
 1               5                  10                  15
Gly Ile Met Leu Leu Gly Ile Leu Met Ile Cys Ser Ala Ala Glu Asn
            20                  25                  30
Leu Trp Val Thr Val Tyr Tyr Gly Val Pro Val Trp Lys Glu Ala Thr
        35                  40                  45
Thr Thr Leu Phe Cys Ala Ser Asp Ala Lys Ala Tyr Asp Thr Glu Val
    50                  55                  60
His Asn Val Trp Ala Thr His Ala Cys Val Pro Thr Asp Pro Asn Pro
65                  70                  75                  80
Gln Glu Val Val Leu Glu Asn Val Thr Glu Asn Phe Asn Met Trp Lys
                85                  90                  95
Asn Asn Met Val Glu Gln Met His Glu Asp Ile Ile Ser Leu Trp Asp
            100                 105                 110
Gln Ser Leu Lys Pro Cys Val Lys Leu Thr Pro Leu Cys Val Thr Leu
        115                 120                 125
Asn Cys Thr Asn Val Asn Val Thr Asn Leu Lys Asn Glu Thr Asn Thr
    130                 135                 140
Asn Ser Ser Ser Gly Gly Glu Lys Met Glu Glu Gly Glu Met Lys Asn
145                 150                 155                 160
Cys Ser Phe Asn Val Thr Thr Leu Ile Arg Asn Lys Arg Lys Thr Glu
                165                 170                 175
Tyr Ala Leu Phe Tyr Lys Leu Asp Val Met Pro Ile Asp His Asp Asn
            180                 185                 190
Thr Ser Tyr Thr Leu Ile Asn Cys Lys Ser Ser Thr Ile Thr Gln Ala
        195                 200                 205
```

```
Cys Pro Lys Val Ser Phe Glu Pro Ile Pro Ile His Tyr Cys Ala Pro
    210                 215                 220
Ala Gly Phe Ala Ile Leu Lys Cys Asn Asp Lys Lys Phe Asn Gly Lys
225                 230                 235                 240
Gly Pro Cys Lys Asn Val Ser Thr Val Gln Cys Thr His Gly Ile Arg
                245                 250                 255
Pro Val Val Ser Thr Gln Leu Leu Leu Asn Gly Ser Leu Ala Glu Glu
            260                 265                 270
Asp Ile Val Ile Arg Ser Glu Asn Phe Thr Asp Asn Ala Lys Asn Ile
        275                 280                 285
Ile Val Gln Leu Asn Val Ser Ile Glu Ile Asn Cys Thr Arg Pro Asn
    290                 295                 300
Asn Asn Thr Arg Lys Lys Ile Thr Leu Gly Pro Gly Arg Val Leu Tyr
305                 310                 315                 320
Thr Thr Gly Glu Ile Ile Gly Asp Ile Arg Arg Ala His Cys Asn Leu
                325                 330                 335
Ser Arg Thr Ser Trp Asn Asn Thr Leu Lys Gln Ile Val Glu Lys Leu
            340                 345                 350
Arg Glu Ile Lys Gln Phe Lys Asn Lys Thr Ile Val Phe Lys Gln Ser
        355                 360                 365
Ser Gly Gly Asp Pro Glu Ile Val Met His Ser Phe Asn Cys Gly Gly
    370                 375                 380
Glu Phe Phe Tyr Cys Asn Ser Thr Gln Leu Phe Asn Ser Thr Trp His
385                 390                 395                 400
Ala Asn Gly Thr Trp Lys Asn Thr Glu Gly Ala Asp Asn Asn Ile Thr
                405                 410                 415
Leu Pro Cys Arg Ile Lys Gln Ile Ile Asn Arg Trp Gln Glu Val Gly
            420                 425                 430
Lys Ala Met Tyr Ala Pro Pro Ile Glu Gly Gln Ile Arg Cys Leu Ser
        435                 440                 445
Asn Ile Thr Gly Leu Leu Leu Thr Arg Asp Gly Gly Ser Ser Glu Glu
    450                 455                 460
Asn Gln Thr Glu Ile Phe Arg Pro Gly Gly Gly Asn Met Lys Asp Asn
465                 470                 475                 480
Trp Arg Ser Glu Leu Tyr Lys Tyr Lys Val Val Lys Ile Glu Pro Leu
                485                 490                 495
Gly Val Ala Pro Thr Lys Ala Lys Arg Arg Val Val Gln Arg Glu Lys
            500                 505                 510
Arg
```

<210> 2
<211> 112
<212> PRT
<213> Human Immunodeficiency Virus

<400> 2

```
Met Ser Glu Val Gln Gln Leu Pro Ile Arg Ala Val Gly Glu Tyr Val
 1               5                  10                  15
Ile Leu Val Ser Glu Pro Ala Gln Ala Gly Asp Glu Glu Val Thr Glu
            20                  25                  30
Ser Gly Leu Ile Ile Gly Lys Arg Val Gln Gly Glu Val Pro Glu Leu
        35                  40                  45
```

```
Cys Val Val His Ser Val Gly Pro Asp Val Pro Glu Gly Phe Cys Glu
    50                  55                  60
Val Gly Asp Leu Thr Ser Leu Pro Val Gly Gln Ile Arg Asn Val Pro
65                  70                  75                  80
His Pro Phe Val Ala Leu Gly Leu Lys Gln Pro Lys Glu Ile Arg Lys
                85                  90                  95
Gln Lys Phe Val Thr Cys His Tyr Lys Ala Ile Pro Cys Leu Tyr Lys
            100                 105                 110
```

Fig. 1A

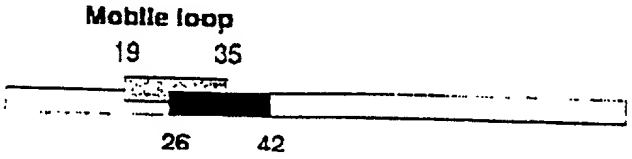


Fig. 1B

Fig. 2A

Fig. 2B

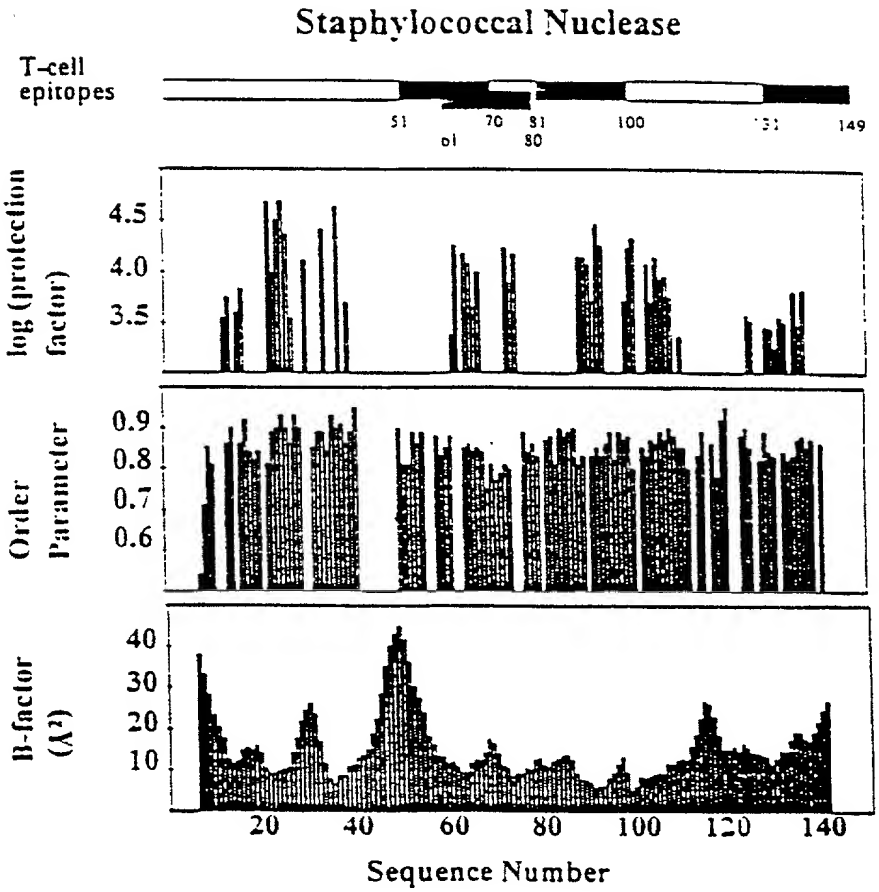
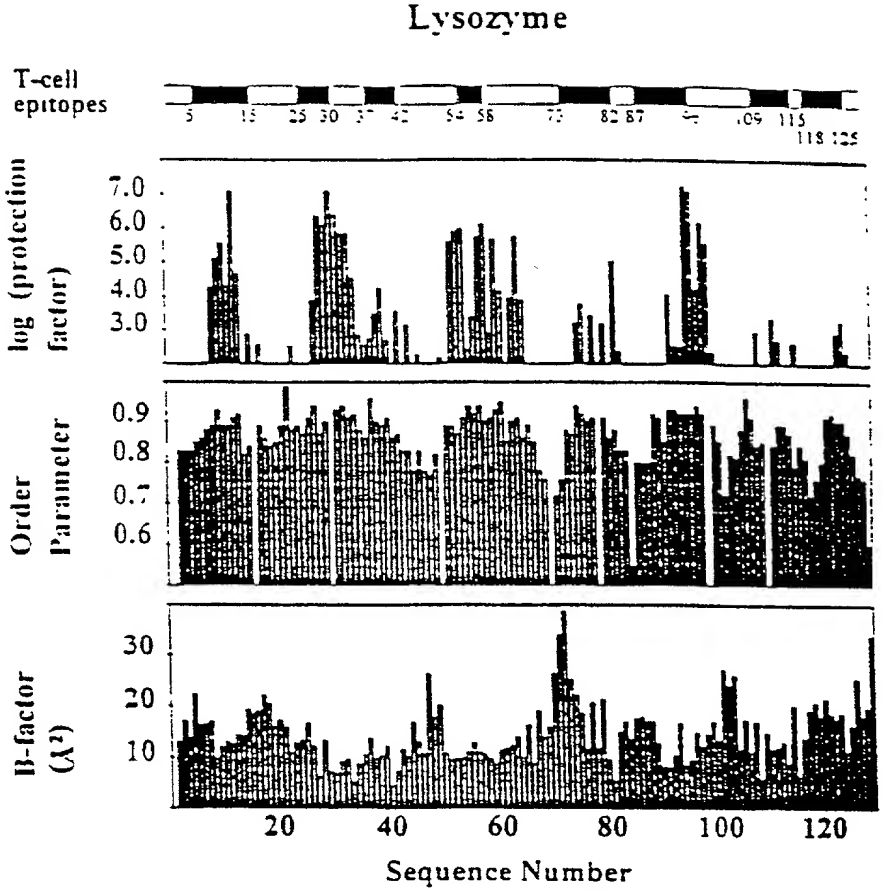

Staphylococcal Nuclease
T-cell epitopes
51
70
81
100
131
149
α1
β1
β2
log (protection factor)
4.5
4.0
3.5
Order Parameter
0.9
0.8
0.7
0.6
B-factor (Å²)
40
30
20
10
20
40
60
80
100
120
140
Sequence Number

Fig. 3A

Fig. 3B


Lysozyme
T-cell epitopes
log (protection factor)
Order Parameter
B-factor (Å²)
Sequence Number

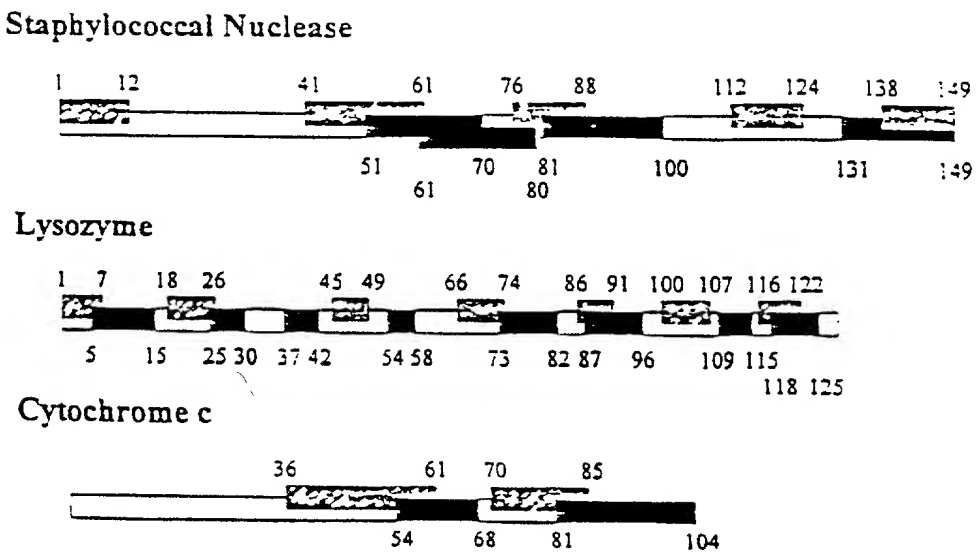

Staphylococcal Nuclease
1 12 41 61 76 88 112 124 138 149
51 70 81 100 131 149
61 80
Lysozyme
1 7 18 26 45 49 66 74 86 91 100 107 116 122
5 15 25 30 37 42 54 58 73 82 87 96 109 115
118 125
Cytochrome c
36 61 70 85
54 68 81 104

Figure 4

Fig. 5A

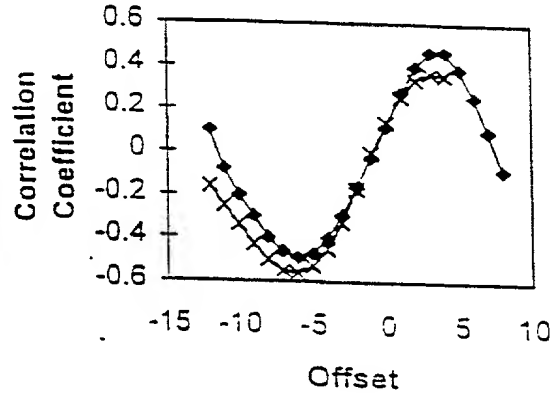
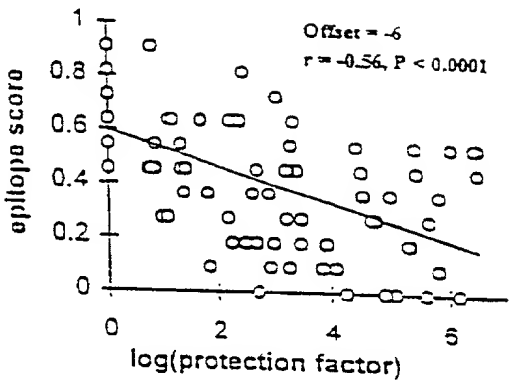

Correlation Coefficient
0.6
0.4
0.2
0
-0.2
-0.4
-0.6
-15 -10 -5 0 5 10
Offset

Fig. 5B


Offset = -6
r = -0.56, P < 0.0001
epitope score
1
0.8
0.6
0.4
0.2
0
0 2 4 6
log(protection factor)

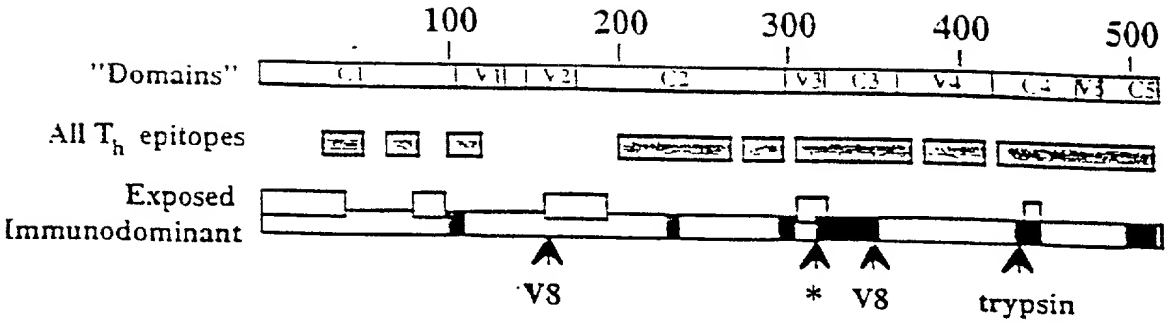

100
200
300
400
500
"Domains"
All T_h epitopes
Exposed
Immunodominant
V8
*
V8
trypsin

Figure 6

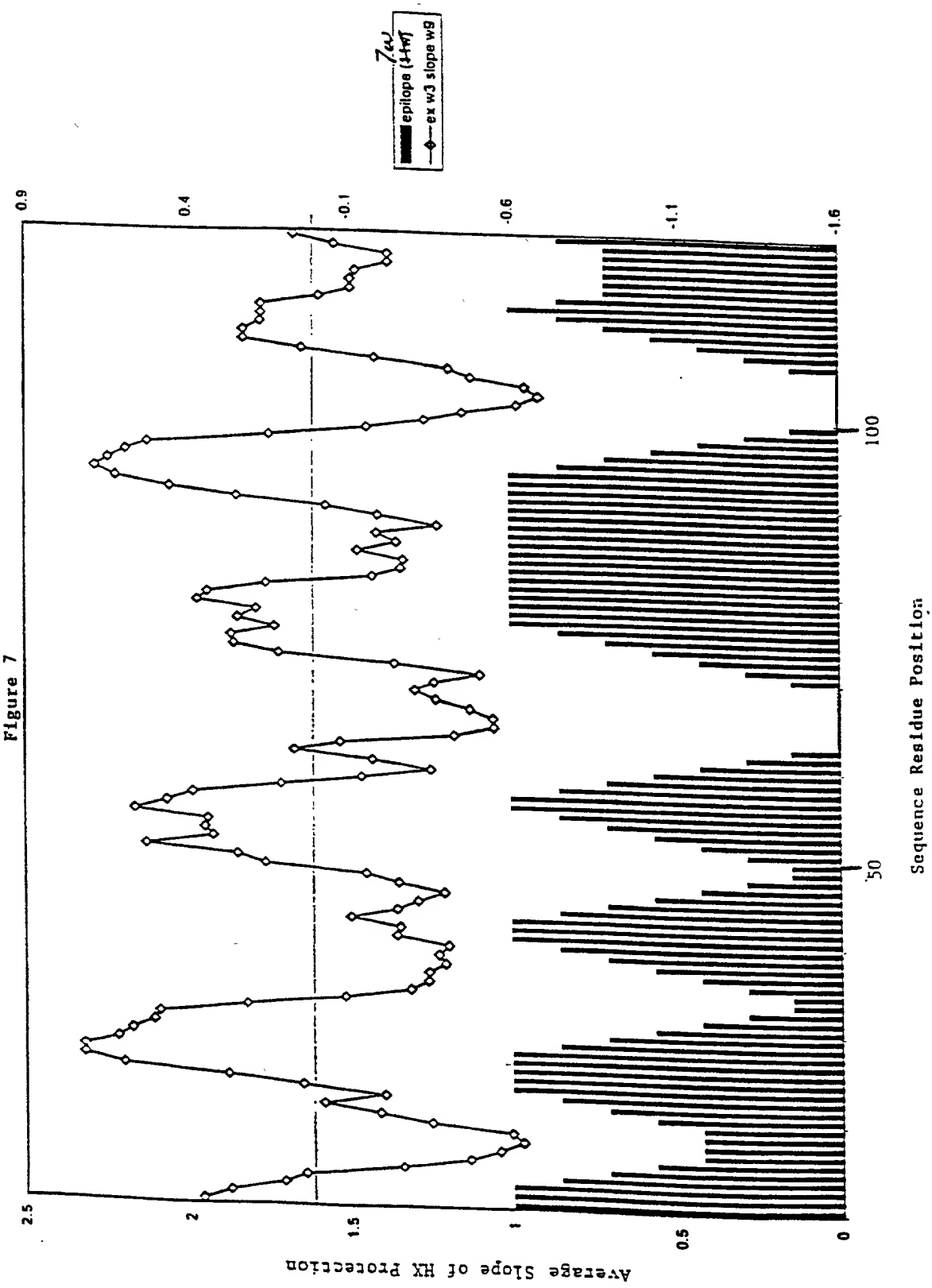

epitope (Tau)
ex w3 slope wg
0.9
0.4
-0.1
-0.6
-1.1
-1.6
2.5
2
1.5
1
0.5
0
50
100
Average Slope of HX Protection
Sequence Residue Position

Figure 7

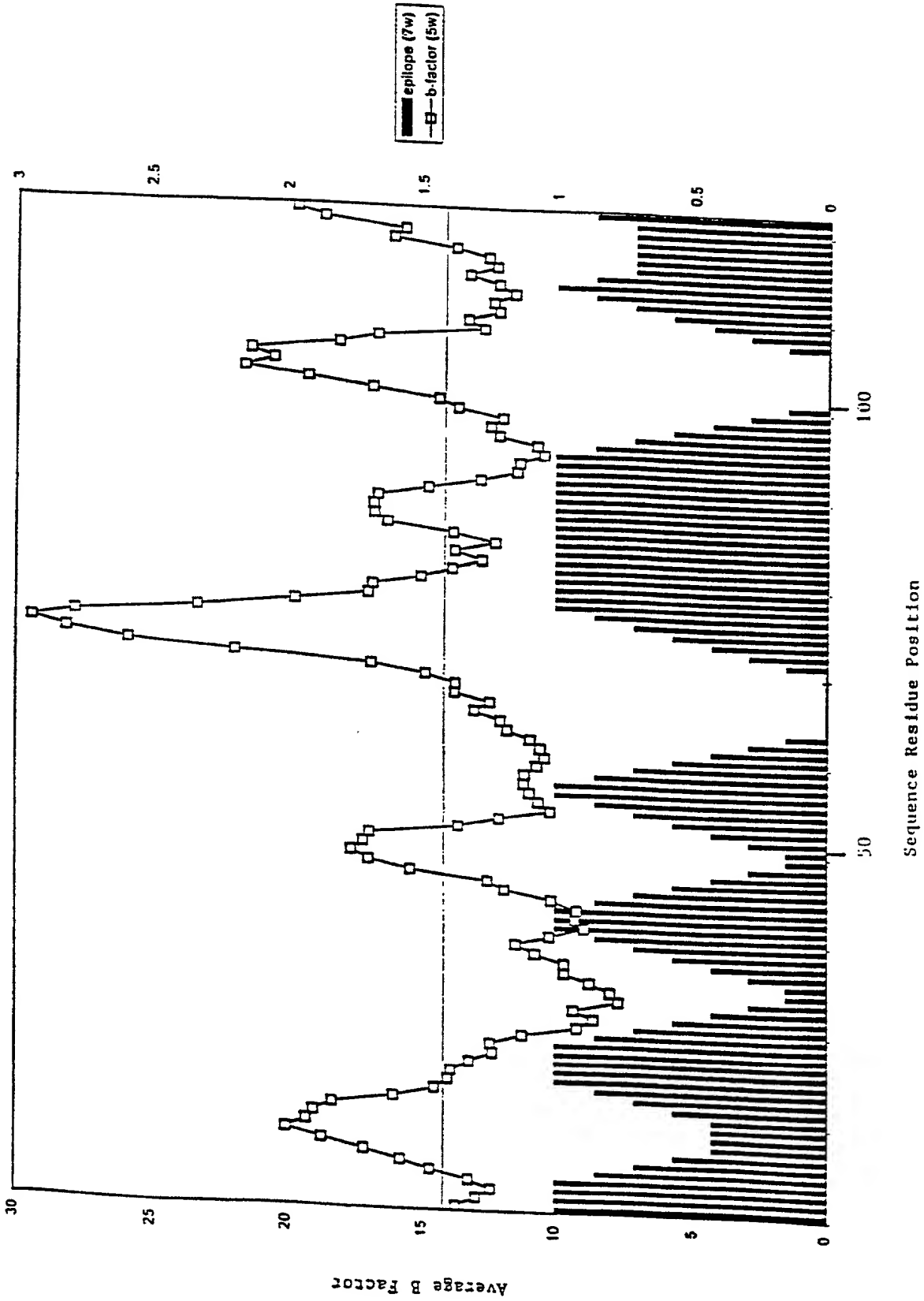

epitope (7w)
b-factor (5w)
Average B Factor
Sequence Residue Position
0
5
10
15
20
25
30
0
0.5
1
1.5
2
2.5
3
50
100

Figure 8

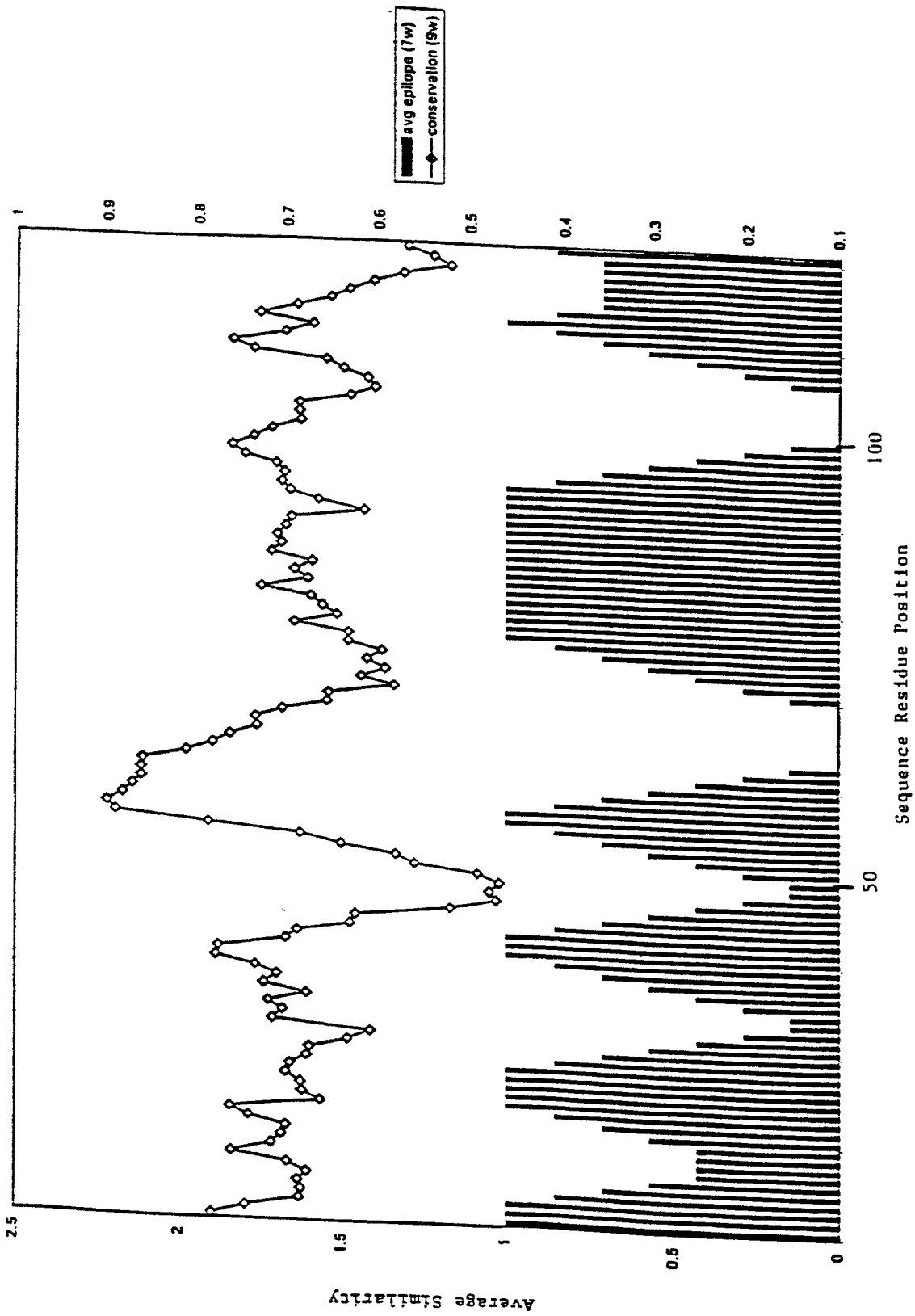

avg epitope (7w)
conservation (9w)
Average Similarity
Sequence Residue Position
1
0.9
0.8
0.7
0.6
0.5
0.4
0.3
0.2
0.1
2.5
2
1.5
1
0.5
0
50
100

Figure 9

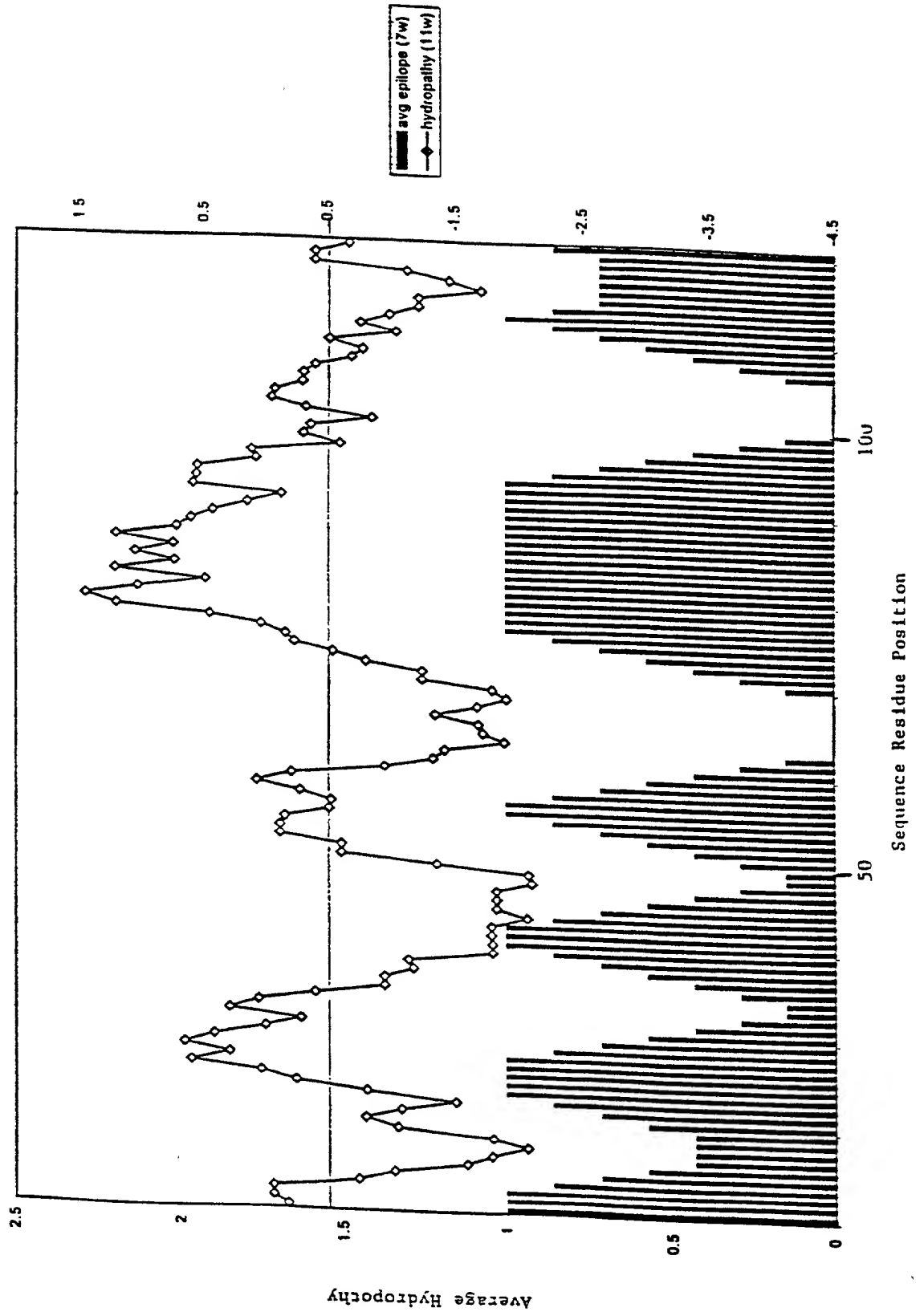

avg epitope (7w)
hydropathy (11w)
1.5
0.5
-0.5
-1.5
-2.5
-3.5
-4.5
2.5
2
1.5
1
0.5
0
Average Hydropathy
50
100
Sequence Residue Position

Figure 10

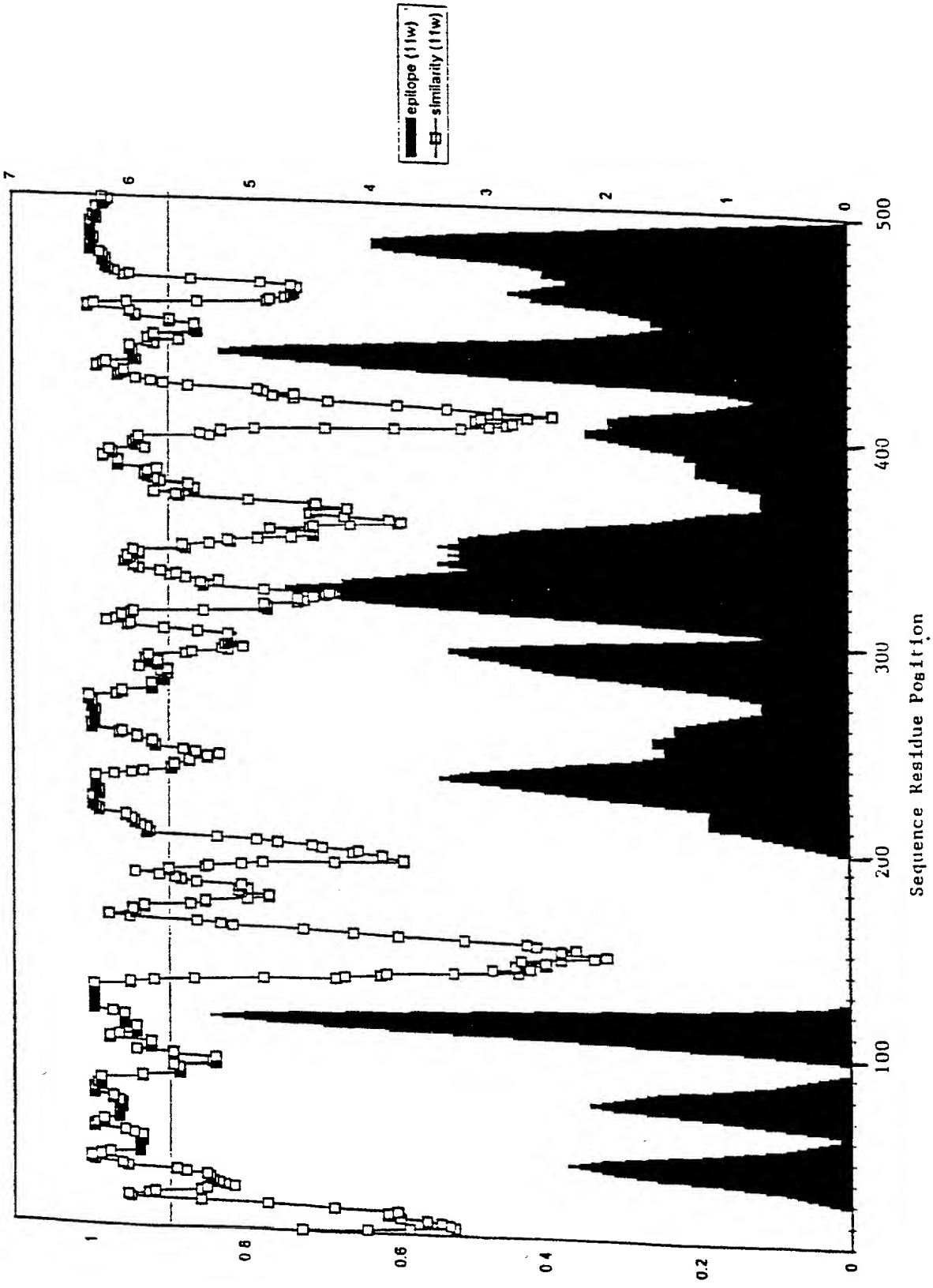


Figure 11A

Figure 11B

NOV 98

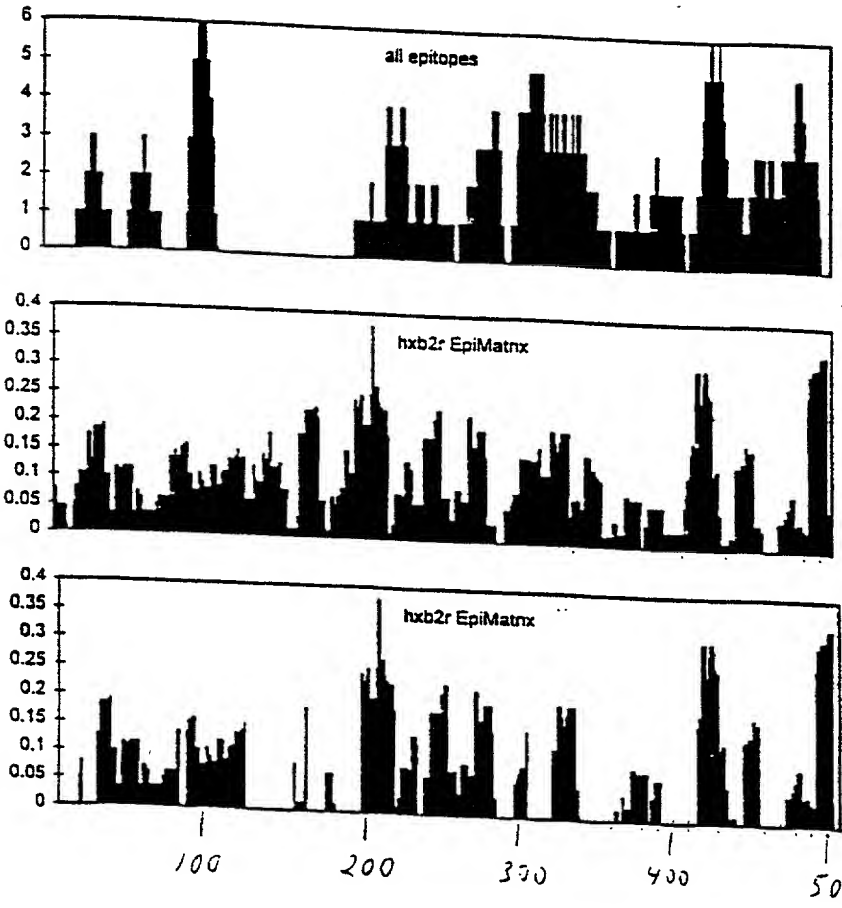


Figure 13A

Figure 13B

Figure 13C

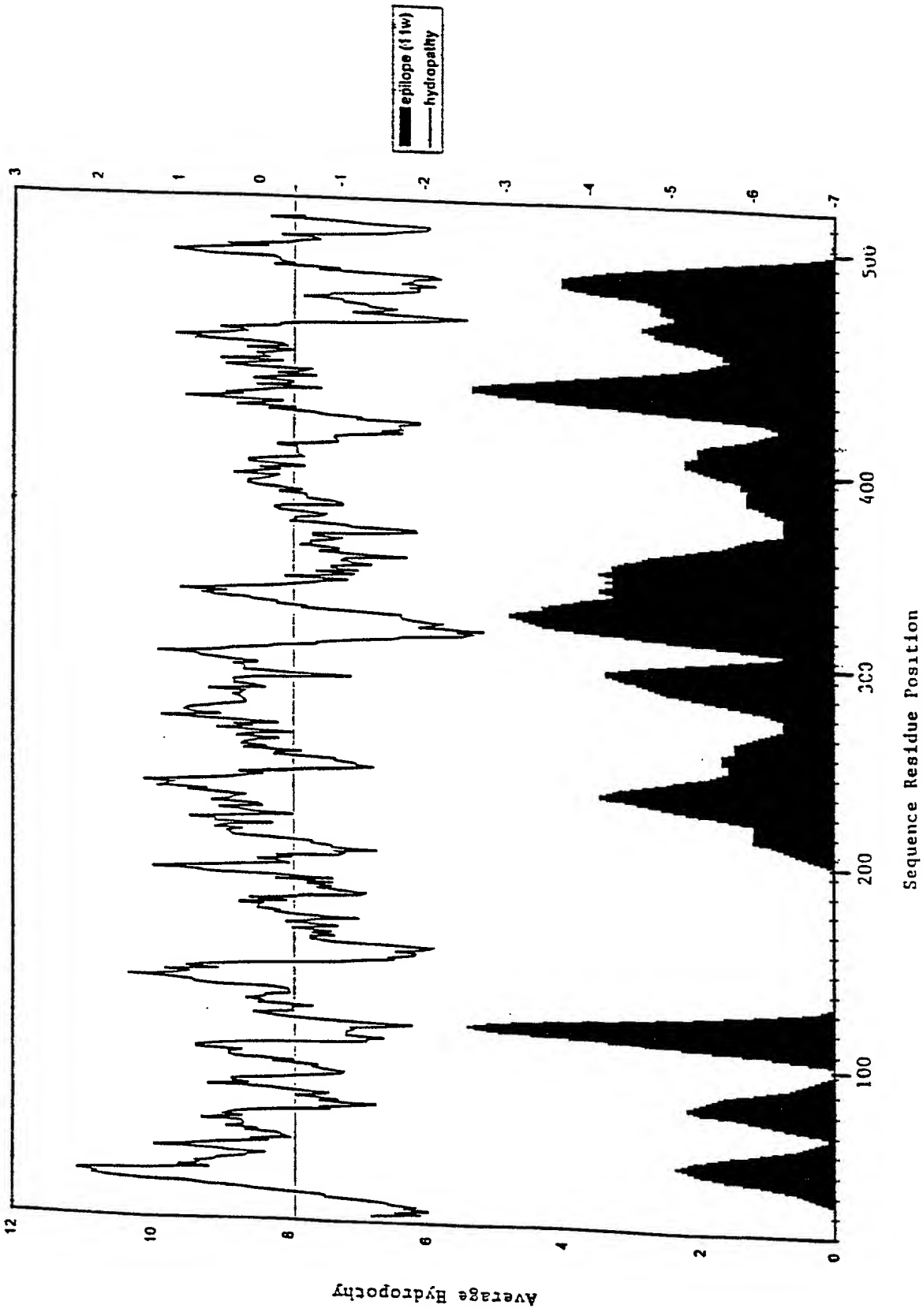

epilope (11w)
hydropathy
Average Hydropathy
Sequence Residue Position
12
10
8
6
4
2
0
3
2
1
0
-1
-2
-3
-4
-5
-6
-7
100
200
300
400
500

Figure 12

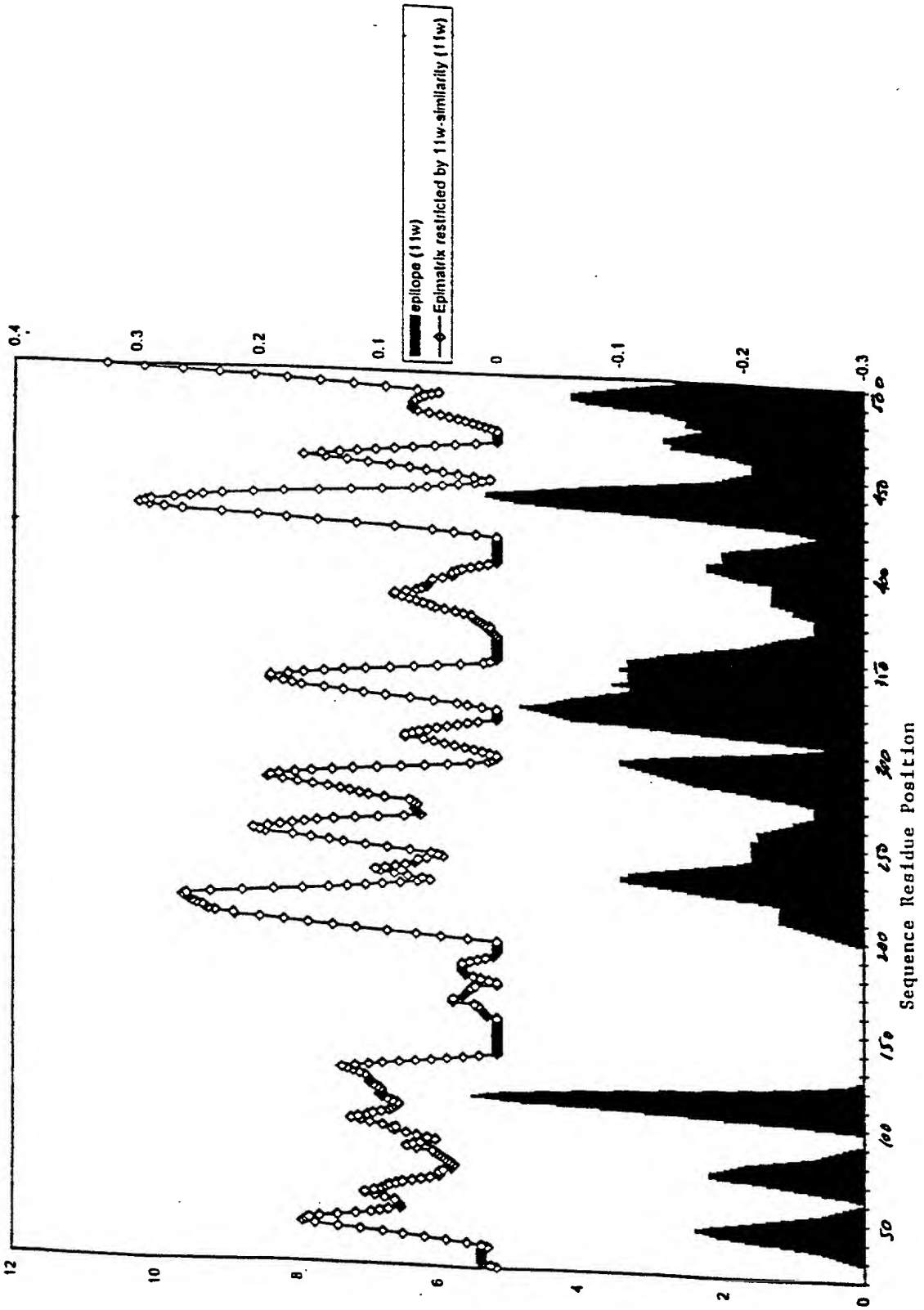
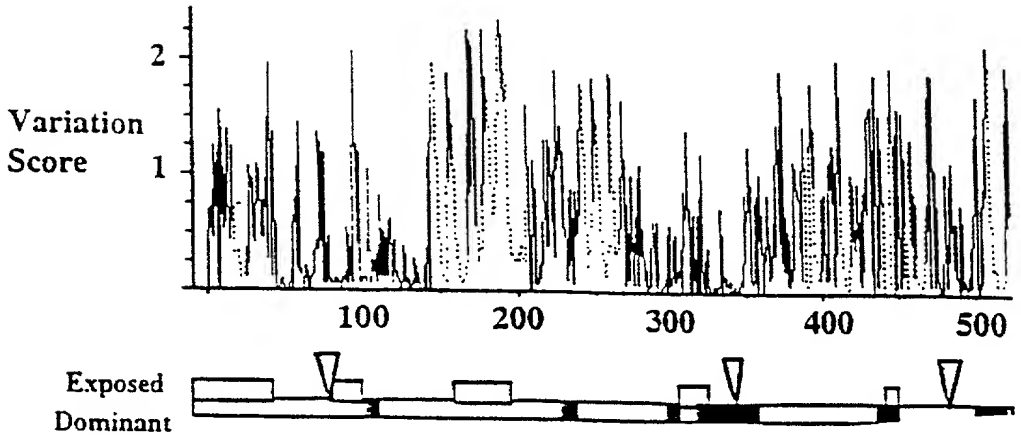

epitope (11w)
Epinatrix restricted by 11w-similarity (11w)
0.4
0.3
0.2
0.1
0
-0.1
-0.2
-0.3
12
10
8
6
4
2
0
50
100
150
200
250
300
350
400
450
500
Sequence Residue Position

Figure 14



Fugure 15

Fig. 16 **A**

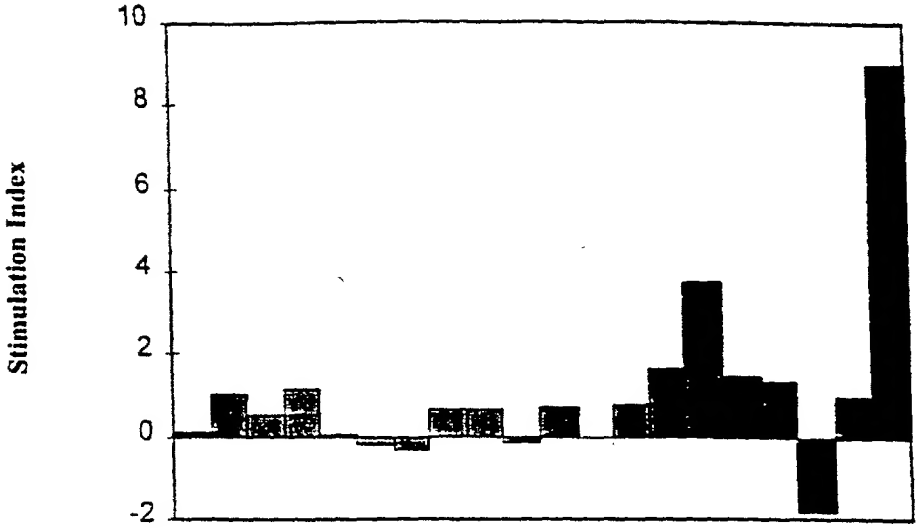


Fig. 16 **B**

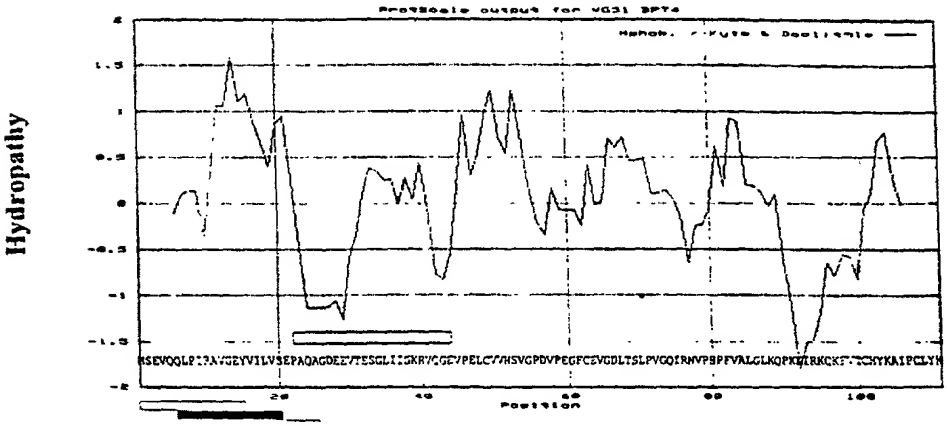


Fig. 16 **C**

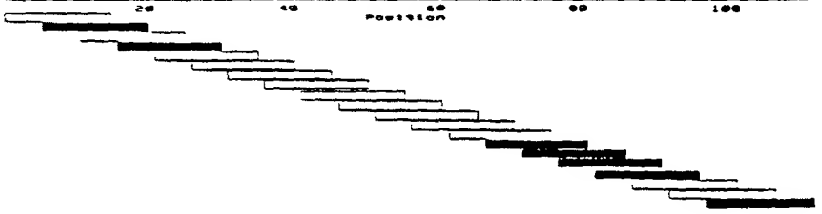
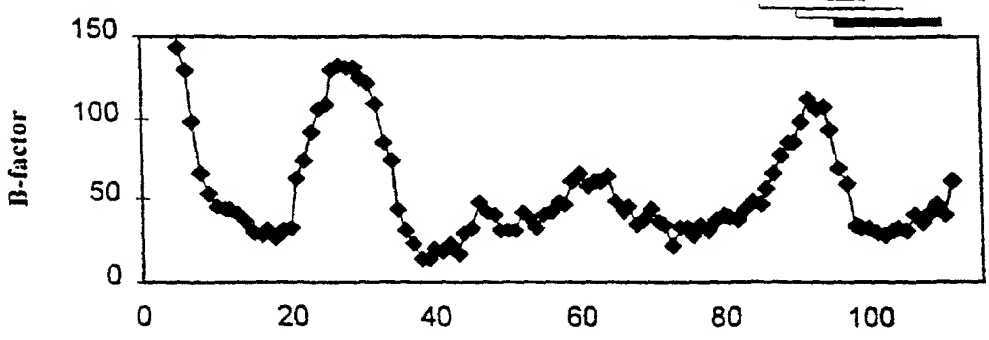


Fig. 16 **D**



20 APR 2000

09/463590

PATENT

ATTORNEY DOCKET NO: 07005/003002

COMBINED DECLARATION AND POWER OF ATTORNEY

As a below named inventor, I hereby declare that:

My residence, post office address and citizenship are as stated below next to my name.

I believe I am the original, first and sole inventor of the subject matter which is claimed and for which a patent is sought on the invention entitled PREDICTION, DETECTION, AND DESIGN OF T-CELL EPITOPES, the specification of which

- ☐ is attached hereto, and
- ☒ was described and claimed in PCT International Application No. PCT/US98/15563 filed on July 28, 1998 and as amended under PCT Article 19 on _____.
- ☒ was filed on January 26, 2000 as Application Serial No. 09/463,590 and was amended on _____.

I hereby state that I have reviewed and understand the contents of the above-identified specification, including the claims, as amended by any amendment referred to above.

I acknowledge the duty to disclose all information I know to be material to patentability in accordance with Title 37, Code of Federal Regulations, §1.56(a).

FOREIGN PRIORITY RIGHTS: I hereby claim foreign priority benefits under Title 35, United States Code, §119 of any foreign application(s) for patent or inventor's certificate or of any PCT international application(s) designating at least one country other than the United States of America listed below and have also identified below any foreign application for patent or inventor's certificate or any PCT international application(s) designating at least one country other than the United States of America filed by me on the same subject matter having a filing date before that of the application(s) of which priority is claimed:

Country	Serial Number	Filing Date	Priority Claimed?
PCT	PCT/US98/15563	July 28, 1998	Yes

PROVISIONAL PRIORITY RIGHTS: I hereby claim priority benefits under Title 35, United States Code, §119(e) and §120 of any United States provisional patent application(s) listed below filed by an inventor or inventors on the same subject matter as the present application and having a filing date before that of the application(s) of which priority is claimed:

Serial Number	Filing Date	Status
60/054,156	July 29, 1997	Abandoned

NON-PROVISIONAL PRIORITY RIGHTS: I hereby claim the benefit under Title 35, United States Code, §120 of any United States application(s) listed below and, insofar as the subject matter of each of the claims of this application is not disclosed in the prior United States application in the manner provided by the first paragraph of Title 35, United States Code, §112, I acknowledge the duty to disclose all information I know to be material to patentability as defined in Title 37, Code of Federal Regulations, §1.56(a) which became available between the filing date of the prior application and the national or PCT international filing date of this application:

COMBINED DECLARATION AND POWER OF ATTORNEY

Serial Number	Filing Date	Status

I hereby appoint the following attorneys and/or agents to prosecute this application and to transact all business in the Patent and Trademark Office connected therewith: Paul T. Clark, Reg. No. 30,162, Karen L. Elbing, Ph.D. Reg. No. 35,238, Kristina Bieker-Brady, Ph.D. Reg. No. 39,109, Susan M. Michaud, Ph.D. Reg. No. 42,885, Mary Rose Scozzafava, Ph.D., Reg. No.36,268, James D. DeCamp, Ph.D., Reg. No. 43,580.

Address all telephone calls to: Kristina Bieker-Brady, Ph.D. at 617/428-0200.

Address all correspondence to: Kristina Bieker-Brady, Ph.D. at Clark & Elbing LLP, 176 Federal Street, Boston, MA 02110.

I hereby declare that all statements made herein of my own knowledge are true and that all statements made on information and belief are believed to be true; and further that these statements were made with the knowledge that willful false statements and the like so made are punishable by fine or imprisonment, or both, under Section 1001 of Title 18 of the United States Code and that such willful false statements may jeopardize the validity of the application or any patents issued thereon.

Full Name (First, Middle, Last)	Residence Address (City, State, Country)	Post Office Address (Street, City, State, Country)	Citizenship
Samuel J. Landry	New Orleans, LA USA	7925 Nelson Street New Orleans, LA 70125	USA
Signature:			Date: 4/5/00